DOCUMENT RESUME

ED 179 564 TM 009 841

AUTHOR Russell, Dale; Plakos, John
TITLE Developing a Pupil Assessment System for Proficiency-Based Instructional Programs.
INSTITUTION Los Angeles County Superintendent of Schools, Calif.
PUB DATE Sep 77
NOTE 98p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Standards; Achievement Tests; *Criterion Referenced Tests; Educational Accountability; *Educational Assessment; Educational Objectives; Educational Philosophy; Elementary Secondary Education; Measurement Techniques; *Performance Based Education; *Program Development; Student Testing; Test Construction; *Testing Programs; Test Results; Test Selection

ABSTRACT

A method of developing measurement instruments to assess academic achievement in proficiency-based education is presented. Materials and exercises are sequenced to correspond to major activities which must be completed in developing local assessment systems. The exercises contained in this workshop guide have been field tested and can be followed by both large and small districts. The method is comprised of ten steps: (1) establishing a school philosophy; (2) identifying educational goals; (3) developing performance objectives for each goal; (4) identifying pupils' skills to be assessed; (5) selecting or constructing measurement instruments; (6) setting performance standards; (7) developing procedures for monitoring student proficiencies and reporting data; (8) implementing the assessment procedures and collecting data; (9) making decisions about students' needs based on the data; and (10) adjusting curricula and programs according to the evaluation results. Test validity and test reliability are also discussed. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179564

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Developing A Pupil Assessment System

for Proficiency-Based Instructional Programs

By
Dale Russell and John Plakos

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Diana Butcher
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 841

SEPTEMBER 1977
OFFICE OF THE LOS ANGELES COUNTY SUPERINTENDENT OF SCHOOLS
DIVISION OF
PROGRAM EVALUATION, RESEARCH, AND PUPIL SERVICES

TABLE OF CONTENTS

		Page
Introduction and Statement of Purpose		1
Process Sequence — Flow Diagram		3
Overview of Ten Steps in Development of Assessment System		4
Workshop Goal		5
Step 1	State District Educational Philosophy	7
	Exercise 1: Establishing a School Philosophy	7
Step 2	Identify Educational Goals	15
	Exercise 2: Identifying Educational Goals	15
Step 3	Develop Performance Objectives	18
	Exercise 3: Examining Performance Objectives	18
	Exercise 4: Relationships Between Course Objectives and Taxonomic Classifications	23
Step 4	Identify Pupil Learning to be Assessed	31
	Exercise 5: Selecting Educational Proficiencies and Assigning Priorities	31
Review Statement		36
Step 5	Construct/Adopt Assessment Instruments	37
	Exercise 6: Instrument Selection	43
	Exercise 7: Developing Instructional Specifications and Items, Assessing Item Adequacy	46
	Exercise 7; Part A: Instrument Specifications and Item Development	55
	Exercise 7; Part B: Assessing Item Adequacy	61
Step 6	Set Performance Standards	65
	Exercise 8: Determining Performance Standards	65
	Exercise 8; Part A	66
	Exercise 8; Part B	69
Step 7	Develop Monitoring and Reporting Procedures	72
Step 8	Implement the Proficiency Assessment System and Collect Data	79
Step 9	Make Decisions About Respondents Based on Data	80
Step 10	Make Program Refinements When Required	81
Test Reliability and Validity		82
Summary		85
References and Suggested Readings		86

Introduction and Statement of Purpose

Accountability in public education has been a widely discussed concept for several years. Previous legislation has attempted to promote accountability by encouraging district autonomy in establishing graduation requirements (SB-1), showing cost effectiveness of educational efforts (PPBS) and other similar strategies. The passage of AB 3408 (1977), authored by Assemblyman Gary Hart, provides a more powerful legal incentive for accountability than previously has existed. The "Hart Law" requires local districts to establish minimum competencies for graduation in the areas of reading, writing, and math. This law further states that, after June 1980, no student shall graduate from high school who has not met the standards of proficiency established by the local board of education. More recent legislation (i.e., AB 65, 1977, authored by Assemblyman LeRoy Greene) expanded the requirement for district verification of pupil proficiencies to include pupils in elementary school.

How districts assess proficiencies is to remain a local decision. Many districts are seeking assistance in developing and implementing their proficiency assessment systems. A major concern expressed by district personnel is that they want to comply with legal requirements but still use the assessment data in curriculum development and renewal. This is a valid concern. They do not want assessment to be separated from the curriculum. Pupil assessment must be responsive to existing curricula. Districts that develop an assessment system which is not designed to reflect learning occurring in their local instructional programs are inviting future problems. They will have difficulty interpreting assessment findings, defending proficiency standards, gaining staff acceptance of the assessment system and other similar problems. If some basic safeguards are observed, however, assessing pupil proficiencies is a relatively easy process to conceptualize and can be done well, even by districts with limited resources. The safeguards most important to observe when developing an assessment system are related to the area of instrument selection or construction.

There are presently two types of measures which may be used to assess learner gains in such areas as reading, writing, and computation: norm-referenced and criterion-referenced. The differences between norm and criterion-referenced measures have sometimes been exaggerated to illustrate the inadequacy of norm-referenced measures for individual assessment, and, in some cases, program evaluation. Norm and criterion-referenced measures may differ widely or be identical in all important aspects except construction procedures. Both may have norms, may indicate a performance standard to certify mastery, and may have standardized administration and scoring procedures. The **purpose** for which a test is constructed, however, is the central variable to consider in selecting appropriate instrumentation. The purpose for constructing norm and criterion-referenced tests **always** clearly separates the two types of measures. Norm-referenced tests are developed for the purpose of relating an individual's performance on that test to the performance of others who took the test; i.e., the norming group. Criterion-referenced tests, conversely, are developed to determine whether an individual can perform acceptably when compared to a pre-established standard.

The new laws provide a mechanism whereby the quality of an educational program can be inferred **partially** from the ability of students to attain locally established minimum standards of proficiency. Therefore, it is appropriate for students to be measured on instruments which were constructed to have a direct relationship to the stated district course of study, and district verification of learner proficiency should be based on criterion-referenced measures.

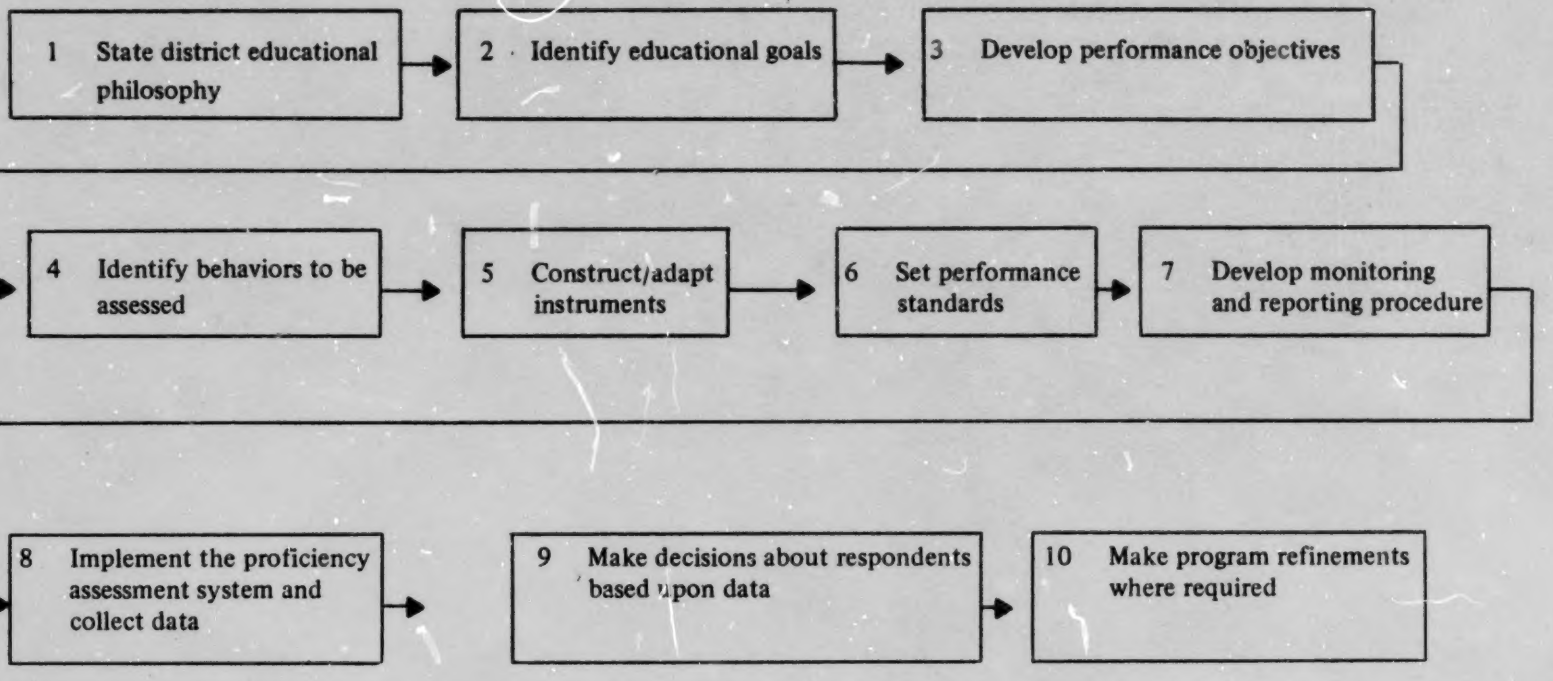
Popham (1976) stated well the case against using norm-referenced tests for individual assessment when certifying competency:

> The main problem with relying on standardized (norm-referenced) achievement tests as the vehicles for establishing performance standards is that such tests provide miserable instructional targets. The tests are described in an artfully vague fashion to make them appealing to a wide range of potential purchasers. This ambiguity makes it most difficult to teach successfully toward the competencies incorporated in the test, because such competencies are rarely isolated or described by the test publishers. And a district **must** try to teach effectively toward its graduation requirements. If it doesn't after 1980, there will be loads of youngsters who'll not get a diploma because, in essence, the district didn't give them sufficiently effective instruction.
>
> A district which wants a maximum number of its students to master the key competencies it designates must clearly identify for its instructional staff exactly what those competencies are. Standardized achievement tests do not provide lucid instructional targets. Therefore, their use almost certainly means that a substantial number of a district's students will fail.

There is a scarcity of personnel at the local district level with expertise in test construction, and the literature on criterion-referenced test construction is generally sparse. Therefore, this document was prepared to provide an outline of concepts which underlie criterion-referenced test construction and practice in applying some of these key concepts. It is hoped that school personnel desiring to construct local criterion-referenced measures will be able to apply the concepts presented in this document as they begin to assess student competencies in their districts.

The purpose of this workshop is to illustrate one method of developing instruments appropriate for use in a proficiency-based system. Workshop materials and exercises are sequenced to correspond to major activities districts must complete when developing their local assessment systems. The exercises have been field tested, and it has been shown that large and small districts can follow the procedures outlined in this document to develop an assessment system that has wide acceptance among various school audiences. The process of developing a pupil proficiency assessment system consists of 10 steps and are shown on the following page.

Process Sequence for Developing a Pupil Proficiency Assessment System


1 State district educational philosophy
2 Identify educational goals
3 Develop performance objectives
4 Identify behaviors to be assessed
5 Construct/adapt instruments
6 Set performance standards
7 Develop monitoring and reporting procedure
8 Implement the proficiency assessment system and collect data
9 Make decisions about respondents based upon data
10 Make program refinements where required

Overview of the Ten Steps for the development of a Pupil Proficiency Assessment System

1 State the district's educational philosophy. The school philosophy should be developed using input received from all audiences of the school community.

2 Identify the district's educational goals. The goals must be consistent with the district philosophy and must provide the basic definition of the curriculum scope and sequence.

3 Develop performance objectives for each goal which permit school managers to assess the level of pupil proficiency in given curriculum areas.

4 Determine which areas of pupil behavior or learning require assessment in a particular district. The selection of the areas to be assessed should be based on explicit criteria. For example, some districts will choose to assess only basic skills, other districts will also assess physical performance, fine arts, vocational preparation, or other valued areas of curriculum. Also, a district must decide whether assessment will focus on proficiency as demonstrated in a school context, simulated or actual life situations.

5 State assessment specifications to guide instrument construction or adaptation. This is the first task when constructing or adapting instruments.

a. Assessment specifications must include:
   - A statement of the different achievement and taxonomic levels of the objectives to be assessed.
   - A description of the situations in which pupils must perform to verify that they have attained desired proficiencies.
   - A description of the particular tasks pupils must perform to verify that they have attained desired proficiencies.
   - A description of the observable outcomes of the assigned tasks that demonstrate pupils have attained the required proficiency.

b. Construct test questions or observation strategies which can provide objective data on learner performance in the identified situations. These questions and/or observation strategies must meet high standards of test construction.

c. Select a representative sample of items and place them in a test format for field testing. Items used to certify mastery of a particular class of behaviors should be homogeneous in content and performance level.

d. Derive a scoring procedure which will yield a score capable of expressing, objectively and meaningfully for each pupil behavior being assessed, the individual's performance level and characteristics.

6 Set performance standards which pupils must achieve. Standards of proficiency provide a mechanism for differentiating those pupils who have acquired an expected amount of skill or knowledge from those who have not. Proficiency standards should reflect agreement among the various school audiences regarding which standards are appropriate.

7 Develop procedures for monitoring pupil proficiencies and reporting data. Data from proficiency assessments are the bases for making judgments about individual learner progress. A well-conceived pupil monitoring and reporting procedure will assure that these data are clear, concise, and readily available.

8 Implement the pupil proficiency assessment system and collect data. The best proficiency assessment system is one which identifies, early in the educational experience those pupils who need alternative instruction.

9 Decide, with the pupil, what course of action would best benefit the pupil. In a proficiency based instructional system, pupils and staff must continually examine whether each pupil is progressing satisfactorily. If a pupil is progressing well, continuation of counseling and instructional support at the level and of the type provided up to that point would be appropriate. Early identification of pupils who are not progressing well, however, provides the pupil, parents, and staff an opportunity to determine alternative courses of study or means of instruction that may better serve the pupil.

10 Refine the educational programs based on assessment data. As districts initiate their proficiency assessment systems, the professional staffs will learn much about which portions of the district curricula are most effective and which are not producing pupil learning at the desired rate. By systematically gathering and analyzing information on pupil performance, it is possible to use the findings to make curricular improvements in areas where deficiencies become apparent.

In the remainder of this document the ten steps for developing a pupil proficiency assessment system are explained. The first six of these steps are also illustrated by workshop exercises related to the underlying concepts of the process.

Workshop Goal

At the conclusion of the workshop, participants will be able to develop a pupil proficiency assessment system appropriate for use in their local school districts.

Following are examples of selected aspects of a pupil proficiency assessment system which participants will be able to develop at the conclusion of the workshop:

Examples of Selected Aspects of a Pupil Proficiency Assessment System

Philosophy

We believe that all pupils should be proficient in reading comprehension, writing, and computational skills to the extent that they can apply them confidently in society. Learning occurs in a variety of environments and requires the active involvement of the learner. Pupils should participate in setting educational goals and in selecting the combination of learning environment options most appropriate for their styles of learning.

Goal: The pupil will be able to comprehend written matter necessary to function in society.

Objective: Upon exit from high school, pupils will be able to utilize clues to derive word meaning in selected reading at the standard prescribed by the Board of Trustees as measured by Reading Proficiency Assessment.

Example of Test Specifications

Sample Item:

Directions:

Choose the answer that means the same as the underlined word as it is used in the sentence.

The **flammable** material easily caught on fire.

_____ A. burnable
_____ B. dense
_____ C. porous
_____ D. supportable

Taxonomic Level: **Comprehension**

Stimulus Attributes (what the test item looks like):

1. The reading level of the items will not exceed eighth grade level as determined by a process recognized as valid by experts in the reading field.
2. Words to be tested for knowledge or interpretation of meaning will be underlined in the stem.
3. All underlined words will be presented in context.
4. Answer choices will be presented alphabetically to the right of the stem and vertically aligned.
5. All response choices will be correctly spelled.

Response Attributes (what the answer consists of):

1. Four answer options will be presented, one of which is accurate.

2. Incorrect answers may be:

   a. antonyms
   b. similarly sounding words
   c. similarly spelled words

3. All response choices must be at the same level of reading difficulty.

Procedures for setting performance standards, monitoring pupil progress and the remaining portions of the workshop are not included in the preceding examples. These procedures and concepts are fully developed in the workshop exercises, however, and their relationship to the assessment system is clearly shown.

STEP 1: STATE DISTRICT EDUCATIONAL PHILOSOPHY

Successfully implemented curricula generally are widely supported by the various audiences of the school community. An educational philosophy expresses values of the school communities in succinct statements which give direction to educational programs within a district.

Several strategies for developing educational philosophies have been used successfully by school districts in California. In general terms, these procedures encourage school managers to obtain input from represententatives of diverse groups, identify educational values of the representatives, arrive at consensus on the most salient identified values, and state a philosophy in a form that can be given to interested persons for review and comment. The philosophy should be adopted by the governing board.

> When the educational philosophy of a district has been developed from a broad base of input, it is more likely to be widely supported.

When developing their philosophies, districts might wish to consider such issues as the setting where learning occurs, how different pupils learn, how learning should be assessed, whether or not pupils should be involved in setting their personal learning goals. If statements of general agreement can be produced regarding issues like these, then the goals and objectives which define the actual instructional processes can be developed more easily.

Exercise 1: Establishing a School Philosophy

CONCEPT: The basis for developing an instructional program is the philosophy, or set of educational beliefs, common to the educators and those they serve. Unless the philosophy is endorsed by those who will implement the program and by those benefiting from it, it probably will not succeed. By using procedures that gain input from all the school communities, educational planners increase the probability that the developed philosophy will represent all views.

PROCEDURE: One method of developing an educational philosophy is to have representatives of school audiences reach consensus on what it is that schools should assist pupils to learn. As an illustration of how to reach consensus, assume that the workshop participants represent diverse school audiences, i.e., parents, staff, pupils, business, and labor representatives, etc. Divide these representatives into clusters of six to eight persons each, assuring a mixture of representation in each cluster, and implement the following:

1. The facilitator identifies five to eight topics on which team members must reach some agreement or identify the particulars of disagreement.

2. Areas that are appropriate to consider when developing a philosophy are student development, curriculum, instruction, learning, staffing, and organization.

3. Each participant is to read the next five pages independently, marking the statements as directed.

4. Each team then:

   a. Defines (informally and verbally) the meaning of each statement.
   b. Determines agreement-disagreement among team members on each statement.
   c. Reviews the items on which the groups had the greatest divergence and work toward concensus on the divergent statements.

5. Teams then write a statement of educational philosophy on at least one topic they have discussed. The statement must cover a topic that the members of the group agree must be considered when planning an educational program. The developed statement will be used in later exercises, so someone from the group should record it.

Instructions: Make a mark somewhere along each line at a point which best represents your own feelings about each statement.

I. Assumptions About Pupil's Learning

A. Motivation

1. Pupils are innately curious and will explore their environment without adult intervention.

strongly agree	agree	no strong feeling	disagree	strongly disagree

2. Exploratory behavior is self-perpetuating.

strongly agree	agree	no strong feeling	disagree	strongly disagree

B. Conditions for Learning

3. The pupil will display natural exploratory behavior if s/he is not threatened.

strongly agree	agree	no strong feeling	disagree	strongly disagree

4. Confidence in self is highly related to capacity for learning and for making important choices affecting one's learning.

strongly agree	agree	no strong feeling	disagree	strongly disagree

5. Active exploration in a rich environment offering opportunity for discovery will facilitate pupil's learning.

strongly agree	agree	no strong feeling	disagree	strongly disagree

6. Pupils have both the competence and the right to make significant decisions concerning their own learning.

strongly agree	agree	no strong feeling	disagree	strongly disagree

7. Pupils will be likely to learn if they are given considerable choice in the selection of the materials they wish to work with and in the choice of questions they wish to pursue with respect to those materials.

strongly agree	agree	no strong feeling	disagree	strongly disagree

8. Given the opportunity, pupils will choose to engage in activities which will be of high interest to them.

strongly agree	agree	no strong feeling	disagree	strongly disagree

9. If a pupil is fully involved in, and is having fun with, an activity, learning is taking place.

strongly agree	agree	no strong feeling	disagree	strongly disagree

C. Social Learning

10. When two or more pupils are interested in exploring the same problem or the same materials, they will often choose to collaborate in some way.

strongly agree	agree	no strong feeling	disagree	strongly disagree

11. When a pupil learns something which is personally important s/he will wish to share it with others.

strongly agree	agree	no strong feeling	disagree	strongly disagree

D. Intellectual Development

12. Concept formation proceeds very slowly.

strongly agree	agree	no strong feeling	disagree	strongly disagree

13. Pupils learn and develop intellectually not only at their own rate but in their own style.

strongly agree	agree	no strong feeling	disagree	strongly disagree

14. Pupils pass through similar stages of intellectual development buy vary in style and rate.

strongly agree	agree	no strong feeling	disagree	strongly disagree

15. Intellectual growth and development take place through a sequence of concrete experiences followed by abstractions.

strongly agree	agree	no strong feeling	disagree	strongly disagree

16. Verbal abstractions should follow direct experience with objects and ideas, not precede them or substitute for them.

strongly agree	agree	no strong feeling	disagree	strongly disagree

E. Evaluation

17. The preferred source of verification for a pupils solution to a problem comes through examining the materials used in the solution.

strongly agree	agree	no strong feeling	disagree	strongly disagree

18. Errors are necessarily a part of the learning process; they are to be expected and even desired, for they contain information essential for further learning.

strongly agree	agree	no strong feeling	disagree	strongly disagree

19. Those qualities of a person's learning which can be carefully measured are not necessarily the most important.

strongly agree	agree	no strong feeling	disagree	strongly disagree

20. Objective measures of performance have a negative effect upon learning.

strongly agree	agree	no strong feeling	disagree	strongly disagree

21. Learning is best assessed intuitively by direct observation.

strongly agree	agree	no strong feeling	disagree	strongly disagree

22. The best way of evaulating the effect of the school experience on pupils is systematic observation over a long period of time.

strongly agree	agree	no strong feeling	disagree	strongly disagree

II. Assumptions About Knowledge

23. The qualilty of **being** is more important than the quality of **knowing**; knowledge is a means of education, not its end. The final test of an education is what one is, not what one knows.

strongly agree	agree	no strong feeling	disagree	strongly disagree

24. Knowledge is a function of one's personal integration of experience and therefore does not fall into neatly separate categories or "disciplines."

strongly agree	agree	no strong feeling	disagree	strongly disagree

25. The structure of knowledge is personal and idiosyncratic; it is a function of the synthesis of each individual's experience with the world.

strongly agree	agree	no strong feeling	disagree	strongly disagree

26. Little or no knowledge exists which is essential for everyone to acquire.

strongly agree	agree	no strong feeling	disagree	strongly disagree

27. It is possible, even likely, that an individual may learn and possess knowledge of a phenomenon and yet be unable to display it publicly. Knowledge resides with the knower, not in its public expression.

strongly agree	agree	no strong feeling	disagree	strongly disagree

Philosophy Worksheet:

Write the team developed philosophy statement on this page. Consider references which can assist in developing the statement such as the County Course of Study, district philosophy, etc.

STEP 2 IDENTIFY EDUCATIONAL GOALS

Prior to the development of a district's instructional program, existing district goals should be reviewed (or developed if they do not exist) to be certain that courses of study which include the goals are educationally sound. The goals should outline the instructional scope and sequence and enable staff to state the conditions of learning and program objectives necessary for goal attainment.

If district or program goals are missing, incomplete, or need refinement, it is desirable to undertake the task of completing the goal structure. Goals should provide a solid and consistent framework for the development of objectives. The basic elements of the goal development/selection process are as follows:

a. Goal Criteria — Determining and agreeing upon what will constitute an acceptable set of goals. It includes scope, sequence, and format and may include a statement on how the goal statements will be used.

b. Statement Development — Actually formulating the wording of the goal statements. This process can be one of revising a given set of statements or creating new statements.

c. Statement Refinement — This step includes aggregating all the goals generated in the development process and compiling a single list.

d. Assign Goal Priority — Assign a priority sequence to the goal statements. A decision could, but probably would not, be that all goals are a number one priority, thereby eliminating placement of a value on individual goals.

e. Consensus — The process of getting final approval on the goals, including their priority assignments from all persons involved in their development.

f. Adoption — Obtaining formal board adoption of the goal statements.

Exercise 2: Identifying Educational Goals

CONCEPT: Goal statements should provide the basis for articulating the courses of study. Program goals separate elements of the educational philosophy into discrete statements of desired pupil learning. Staff can then write objectives and related instructional activities for each desired goal.

PROCEDURE: The key to implementing an articulated course of study is having a goal structure that defines the scope and sequence of the instructional process. Each group of participants has developed a philosophy statement in Exercise 1. Each group will now examine its philosophy statement and derive from the statement a few goals, in one or two areas, that could be the beginning of planning a course of study.

Developed goals can be recorded on the following page. The goals should indicate who, will do what, and an estimate of an approximate level of performance.

EXAMPLES OF GOALS:

By the end of grade eight pupils should be able to read with understanding materials published in popular journals, e.g. Time, Sports Illustrated, and the like.

Upon exit from high school pupils should be able to acquire information through reading at a level that will enable the pupil to function confidently in society.

GOAL WORKSHEET:

Write the team developed philosphy statements on this page. Consider references that can assist in developing the goal statements as was done when developing the philosophy statement.

After goals have been set, it is necessary for districts to place them in priority order. Setting goal priorities is very similar to the process of setting priorities for objectives. The process is explained in Step 4; therefore, it will not be covered in this section.

Participants who are interested in reading more about how to set goals, establish goal priorities and gain a perspective of their role in needs assessment are referred to: Russell, D.; Plakos, J.; Cox, J., **Developing a Workable Needs Assessment Process**, Los Angeles: Office of the Los Angeles County Superintendent of Schools, 1977.

STEP 3 DEVELOP PERFORMANCE OBJECTIVES

Objectives translate program goals into assessable pupil performance. While the educational philosophy and goals give direction to instructional programs, the objectives state specifically who is to learn what, how well, under what conditions, and when the learning is to occur. Objectives, then, are a central part of a pupil proficiency assessment system.

Exercise 3: Examining Performance Objectives

CONCEPT: Objectives specify learner outcomes that should result from participation in particular courses of study. To be valid, assessment of pupil learning must be based on the objectives of a particular course of study. Therefore, when educators intend to assess learning which results from instruction, they must begin by examining the performance objectives.

One way to assess learning related to specific performance objectives is to use criterion-referenced tests. Criterion-referenced tests are constructed to measure the level of proficiency attained by pupils who participated in a particular set of learning activities.

> Viotla (n/d, p. 11) states:
>
> In order for the test to be good, the set of objectives upon which it is based must be well organized and the objectives must be ... specific descriptions of the terminal behavior of students to be expected as a result of instruction.

Objectives are written at various levels of generality, depending on where they are placed in instructional management. Highly specific objectives are stated for classroom use while much more general objectives are appropriate for total program management. Criterion measures can be constructed for the objectives at all levels of generality; data from such criterion measures would have different uses. Data taken from measures related to program objectives provide useful information for program evaluation and subsequent planning in needs assessment. Data taken from classroom measures assist staff to monitor learner attainments relative to prespecified outcomes; they also can be aggregated with information about many other learners for use in program planning and evaluation.

The content and purpose of objectives are widely accepted. Objectives must state who will do what, when, under what conditions, how they will be measured, and at what minimal level of performance (Plakos, 1975). The purpose of objectives is to set precise boundaries of the behavior to be assessed after a participant has completed a set of learning experiences (Popham, 1973). By prespecifying learner behaviors, teaching also is facilitated since the teacher can provide appropriate practice for learners and increase the probability that the learner will master key constructs (Gagne, 1965, pp., 302-343).

Precise statement of objectives is not enough. Educators also must be able to select objectives which facilitate learner attainment of the essential competencies specified in the course of study. In the context of assessing pupil proficiencies, it is also important that educators be able to discern which proficiencies should provide the basis for certifying pupil operational competence to qualify for graduation and receipt of a diploma. The following possible criteria may assist in the selection of proficiencies that provide a suitable basis for qualifying a pupil to graduate.

1. Utility: The proficiency represents skill(s) which the graduate can apply upon graduation. Areas for applying these proficiencies may range from employment opportunities to continuing education.

2. Scope: The proficiency is one which subsumes previously learned skills and concepts. The proficiency should be sufficientiy inclusive that sampling of learner attainments will result in a valid assessment of the degree that a learner has mastered a course of study.

3. Sensitive to instruction: The proficiency should be one from the course of study that school staffs are responsible to teach and which can be learned by the pupils.

PROCEDURE: Below are five performance objectives taken from a pupil handbook at a school where proficiency based instruction was implemented. Examine them to see if they have all the elements of an objective that were discussed above. For example:

Objective 1.

After reading a "letter to the editor" or political statement or article of persuasion or argument, the pupil will be able to reach a critical decision based on what the pupil has read (agree or disagree with what the pupil has read based on a supported or unsupported point of view).

In objective one, it is clear that some criteria are, and others are not, satisfied, i.e.

Who — The pupil

What — Reach a critical decision based on reading a selected letter, statement, or article.

When — Prior to graduation

Condition — After reading a "letter to the editor" or political statement or article of persuasion or argument.

Measure — Not stated

Standard — Not stated

Utility — Possibly the objective was included in the handbook because it was considered by the school audiences of the district to be knowledge pupils would definitely utilize.

Scope — Possibly to determine the scope of the objective, it would be necessary to review the particular course(s) where the objective was taught.

Sensitive — The objective probably is sensitive to instruction. This information is recorded for Objective No. 1 on the Form for "Exercise 3: Examining Performance Objectives."

Participants should now:

1. Individually review the remaining four objectives.
2. Record their individual analyses on the form for Exercise 3.
3. Discuss their individual records with other participants in their group.
4. Compare their records with the comparison sheet for Exercise 3.
5. As a group, identify changes that are needed to make each objective acceptable.

Objective 2.

Given a sample of writing without punctuation or capital letters, the pupil will provide the necessary punctuation and capital letters to make the meaning clear.

Objective 3.

The pupil will be able to find the solution to a problem of the form: If a tax rate is $2.95 per $1000, how much must you pay for taxes on a house assessed for $9,800?

Objective 4.

Given specific situations, the pupil can describe appropriate procedures for satisfactory resolution of complaints about fraudulent or deceptive practices as measured by achieving a minimum score of 20 on a 25 item district developed assessment instrument.

Objective 5.

Following two weeks of instruction in a general math course on presenting statistical information and given a simple bar graph or line graph, the pupil will determine one component when the other is given.

Exercise 3: Examining Performance Objectives

Objective	Who	What	When*	Conditions	Measure	Standard	Utility	Scope	Sensitive
1	X	X	X	X			?	?	?
2									
3									
4									
5									

*Prior to graduation

Comparison Sheet for Exercise 3: Examining Performance Objectives

Objective	Who	What	When	Conditions	Measure	Standard	Utility	Scope	Sensitive
1	X	X	*	X			?	?	?
2	X	X	*	X			?	?	?
3	X	X	*				?	?	?
4	X	X	*	X	X	X	?	?	?
5	X		*	X			?	?	?

*Prior to graduation

Exercise 4: Relationships Between Course Objectives and Taxonomic Classification

CONCEPT: A pupil assessment system should include learnings at a variety of cognitive, affective, and psychomotor levels. Some districts will want pupils to acquire skills that involve abilities representative of different levels of functioning. For example, one district may choose to assess only the knowledge that pupils have acquired while completing a course of study, while another district may assess the ability of pupils to apply their learnings to understanding or evaluating events which influence their lives. Ability to recall information, a low level of cognitive functioning, does not assure that a pupil will be able to evaluate or apply that information. Districts are encouraged to assess pupil learnings at higher as well as lower levels of congnitive functioning. It would be helpful for districts which choose to assess pupil learnings representative of a variety of levels of mental functioning to have a method for showing the relationships between the learnings and levels of functioning being assessed. A simplified illustration of how the concept of assessing pupil proficiencies across levels of cognitive functioning follows:

At a very low level of cognitive functioning, pupils may be able to apply reading rules to pronounce words that appear on paper. At a higher level, pupils might be able to answer questions about what has been read. A higher level still may be the ability to act on something read, to complete an instruction, or further, to determine whether something that was read conveys a message accurately.

It is important to have a balance in the kinds of mental operations, motor skills or affective behavior the schools are trying to assist their pupils to develop. Taxonomies which assist in classifying behaviors in each of these domains are available. To some extent taxonomic classifications overlap. As Peter (1975, p. 78) has noted, "objectives in the cognitive domain cannot be achieved unless content and method elicit satisfying feelings in the learner." Similar statements could be made about the use of manipulative (psychomotor) reinforcers to promote cognitive learning, for example.

It is helpful to have a very clear procedure to show the relationship between course objectives and an appropriate taxonomy. By showing the taxonomic level of each objective, one ascertains the extent to which objectives for various content exist at various levels of the taxonomy. Examples of available taxonomies that have been developed for use by educators are:

Domain	Reference
Affective	Krathwohl, David R.; Bloom, Benjamin, S.; Mesia, Bertram B. *Taxonomy of Educational Objectives - The Classification of Educational Goals, Handbook II: The Affective Domain*. New York: David McKay Co., Inc., 1964.
Cognitive	Lembo, John M. *The Psychology of Effective Classroom Instruction*. Columbus, Ohio: Charles E. Merrill Pub. Co., 1969.
Psychomotor	Harrow, Anita J. *A Taxonomy of the Psychomotor Domain*. New York: David McKay Co., Inc.
	Simpson, Elizabeth. "The Classification of Educational Objectives, Psychomotor Domain". *Illinois Teacher of Home Economics*. 1966-67

One taxonomy which is widely recognized was developed by Benjamin Bloom and a committee of educators. The levels of the cognitive taxonomy Bloom developed are (Bloom, et al., 1956):

1.00 Knowledge:	The ability to remember — recall or recognize — ideas, facts, etc., in a situation in which cues, signals, and clues bring out whatever knowledge has been stored.
2.00 Comprehension:	The ability to understand what is being communicated and make use of it, without necessarily relating it to other materials or seeing its implications.
3.00 Application:	The ability to use abstractions, rules, principles, ideas, and methods in particular and concrete situations.
4.00 Analysis:	The ability to break down some communication into its constituent elements or parts.
5.00 Synthesis:	The ability to work with pieces, parts, elements, etc., and combine or put them together in some way to form a new pattern or structure.
6.00 Evaluation:	The ability to place value on and make judgments about the extent to which materials and methods satisfy criteria.

An example of how a relationship between objectives and taxonomic level of functioning may be shown is presented on page 25. Notice that, in this example, as students advance through the grades they also are expected to utilize higher cognitive processes.

PROCEDURE: On page 26 are some portions of objectives. Examine them and indicate at which taxonomic level each objective would be placed on Bloom's Taxonomy.

Figure 1: Cognitive Taxonomic Classifications of Language Course Objectives

Course Content	Grade Level	Number of Objectives per Taxonomic Level						Total
		Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	
Word Study Writing Paragraphs Thoughtful Listening	9	3	3	2	1	1	1	11
Social Letters Expanding Vocabulary Using the Library	10	2	3	3	2	1	1	12
Group Discussions Thinking Expository Writing	11	1	2	3	3	2	2	13
Semantics The Precils Discussing and Debating	12	1	1	2	3	3	3	13
	all	7	9	10	9	7	7	49

Exercise 4: Taxonomic Distribution of Learner Objectives

Place an "X" in the column under the taxonomic level corresponding to each objective.

Subject: Language

Objective being assessed	Taxonomic Level 1.0	2.0	3.0	4.0	5.0	6.0
Upon graduation from XYZ High School, students will be able to:						
1. Interpret television, radio, and transportation schedules.						
2. Develop a map utilizing a uniform scale which contains signs, graphs, charts, and diagrams.						
3. Compare the advertized claims for a project with the product's actual performance.						
4. Identify differences between generalizations and factual support.						
5. Contrast the themes of two written works.						
6. Produce a communication on a single theme using the content of four related works.						

Exercise 4: Comparison Sheet for Taxonomic Distribution of Learner Objectives

Subject: Language

Taxonomic Level

Objective being assessed	1.0	2.0	3.0	4.0	5.0	6.0
Upon graduation from XYZ High School, students will be able to:						
1. Interpret television, radio, and transportation schedules.		△				
2. Develop a map utilizing a uniform scale which contains signs, graphs, charts, and diagrams.			△			
3. Compare the advertized claims for a project with the product's actual performance.						△
4. Identify differences between generalizations and factual support.	△					
5. Contrast the themes of two written works.				△		
6. Produce a communication on a single theme using the content of four related works.					△	

The Portland, Oregon Public Schools have developed taxonomies that are intended to reflect how teachers teach and pupils learn. A portion of their taxonomy for social science is reproduced, with their permission, on the following two pages as an example of another taxonomy available to educators.

The Portland taxonomy can be conceptualized as an information processing-type model. This concept is particularly clear in part two related to "Processes of inquiry and problem-solving." Pupils first acquire information which they later validate, organize and interpret. Using information, acting on the basis of it and communicating are the higher level functions.

The taxonomy includes an examination of values and the relationship of basic education and social science. A review of the relationship between basic education and social science presents possibilities of how interdiscipline studies can be conceptualized.

The larger portion of the Portland taxonomy, which is not included in this publication, deals specifically with the content of such topics as anthropology, economics, geography, history, political science, psychology and sociology.

Portland, Oregon Public Schools

SOCIAL SCIENCE TAXONOMY (Partial)

1. Values and social science
    - 1.1 Value clarification skills
        - 1.1.1 Personal
        - 1.1.2 Interpersonal
        - 1.1.3 Societal
    - 1.2 Potential value choices
        - 1.2.1 Humans and environment: related values
        - 1.2.2 Society and culture: related values
        - 1.2.3 Personal functioning: related values
2. Processes of inquiry and problem-solving
    - 2.1 Acquiring information
        - 2.1.1 Perceiving, experiencing
        - 2.1.2 Locating and using resources
        - 2.1.3 Observing behavior
        - 2.1.4 Interviewing subjects
    - 2.2 Validating information
        - 2.2.1 Checking authoritativeness of sources
        - 2.2.2 Checking logical consistency, accuracy
        - 2.2.3 Checking relevance to purposes/objectives
        - 2.2.4 Checking adequacy for acting/deciding
    - 2.3 Organizing information
        - 2.3.1 Labeling, naming, numbering, coding
        - 2.3.2 Recording, listing
        - 2.3.3 Classifying, grouping, categorizing
        - 2.3.4 Ordering, sequencing
        - 2.3.5 Manipulating, arranging, transforming, computing
        - 2.3.6 Estimating
        - 2.3.7 Summarizing, abstracting

- 2.4 Interpreting information
  - 2.4.1 Decoding symbols
  - 2.4.2 Inferring, interpolating, extrapolating
  - 2.4.3 Analyzing
  - 2.4.4 Associating, relating, equating
  - 2.4.5 Comparing, contrasting, discriminating
  - 2.4.6 Synthesizing
  - 2.4.7 Testing against standards or criteria
  - 2.4.8 Generalizing
- 2.5 Using information to produce new information
  - 2.5.1 Theorizing, predicting
  - 2.5.2 Formulating hypotheses
  - 2.5.3 Testing hypotheses
- 2.6 Acting on the basis of information
  - 2.6.1 Reacting
  - 2.6.2 Making decisions
  - 2.6.3 Solving problems
  - 2.6.4 Restructuring values
  - 2.6.5 Restructuring behavior
  - 2.6.6 Encoding prior to communication
  - 2.6.7 Creating
- 2.7 Communicating information
  - 2.7.1 Giving information
  - 2.7.2 Recognizing communication problems
  - 2.7.3 Exchanging information
  - 2.7.4 Using communication technology

3. Basic education and social science
   - 3.1 Language and communication skills
     - 3.1.1 Reading
     - 3.1.2 Writing
     - 3.1.3 Speaking and listening
     - 3.1.4 Viewing (Film, TV)
   - 3.2 Mathematical skills
     - 3.2.1 Arithmetic skills
     - 3.2.2 Measurement and statistical skills
   - 3.3 Personal study skills

STEP 4: IDENTIFY PUPIL LEARNING TO BE ASSESSED

Public schools are located in diverse settings and serve diverse pupil populations. Because the modes of instruction, the resources available to districts, and the responses of pupils to instruction are so different, it is important for each district to identify proficiencies that pupils must attain in the local settings.

The best source of information districts can use to decide what they should assess is their established course of study. Once districts have a clear indication of what must be assessed, they can attend to the task of obtaining and using appropriate instruments.

For example, in a situation where the objective is to correctly select a consumer "best buy," the teaching resource may involve reading various local newspaper advertisements, and the assessment may require the pupils to understand the advertisements and act on the presented information. The pupil may be required to correctly perform computations from information contained in the ads, or compare features to find which are best suited to a defined situation.

Another teaching resource could involve classroom presentations by sales personnel of different appliances, "extolling the virtues" of particular brands or models. The assessment procedure could then be to require the pupils to recall facts presented during the presentation or to defend a consumer selection based on the presented facts. This objective must also be assessed by assigning pupils to visit the local business community where they would locate a particular appliance model which satisfies prestated criteria. The assessment procedure would be to determine whether the identified appliance appropriately matched the criteria.

The major task that districts must complete in Step 4 is to identify those learnings which the pupils must attain and which are embodied in the district courses of study.

Exercise 5: Selecting Educational Proficiencies and Assigning Priorities

CONCEPT: A pupil assessment system should be developed within the available resources of a district. By establishing which pupil learnings are considered most important by the various school audiences, it is possible to begin with the highest priority of pupil learning and methodically develop a complete set of assessment instruments

PROCEDURE: One way to put pupil proficiencies in priority order is through a scaling procedure. After the order of proficiency importance has been determined, assessment instruments for those ranked highest in importance are developed first. If a district has sufficient resources, instruments can later be developed for those proficiencies lower in importance. In assigning priorities to educational proficiencies, it is sometimes desirable to know how opinions of parents, students, and staff differ.

For this exercise, participants are assigned to one of three groups: parents, teachers, and support staff/administrators. While in those groups they review the performance indicators established by the District Language Curriculum Committee (see p. 33). While they review the established proficiency indicators, they make any adjustments they believe advisable and rank the indicators according to importance.

To rank the indicators:

1) Assign participants to role play the parts of parents, teachers, and support staff/administrators, forming three distinct groups. The teacher group works independently from the parents, etc.

2) Each participant reviews the performance indicators on page 33.

3) New performance indicators may be added should the group choose to do so. (Participants should be as prudent as possible when identifying instructional performance indicators since measures will need to be constructed for each identified indicator.)

4) Participants individually assign the number of points they wish to allocate to each performance indicator. (Points were established for this exercise at a rate of the midpoint of the scale times the number of performance indicators. When scaling ten performance indicators on a scale of 1 to 9, 50 points are allocated. Since 5 is the midpoint of the range of 1 to 9, 5 multiplied by 10 equals a total of 50 points.)

5) Each individual in a group records his point allocations on the Group Data Chart. Scale the indicators according to importance by group, and contrast the rankings assigned by the three groups.

(Note: This procedure for ranking proficiencies should be repeated for each goal area. For example, it would **not** be appropriate to rank proficiencies in reading comprehension with computational skills proficiencies.

Caution: Some courses of study have large numbers of pupil proficiencies within each subject area. Since it is not feasible to assess all proficiencies in such cases, it is necessary to reduce the proficiencies requiring assessment to a manageable number. A sorting process could be used to conveniently and accurately identify the proficiencies most requiring assessment. For example: If the district math curriculum consists of 100 objectives, they should be categorized as:

1 - Assessment data are not required
2 - Assessment data have limited use
3 - Assessment data are highly desirable
4 - Assessment data are required

The proficiencies **sorted** into the higher priority categories would be **ranked** first and instrument construction could begin.

PERFORMANCE INDICATORS*

GOAL AREA: WRITTEN EXPRESSION AND SPELLING

Proficiency

1. The pupil can demonstrate a knowledge of the predictability of language.
2. The pupil can manipulate sentence parts to change the intent or effect of the sentence.
3. The pupil can recognize that the apostrophe mark is used to show missing letters in two contracted words.
4. The pupil can make accurate choices using the pronunciation key.
5. The pupil can identify different meanings for one word and select the correct meaning for a given context.
6. The pupil can identify words or phrases that reveal a writer's attitude toward his subject.
7. The pupil can organize his or her material in a logical sequence as determined by the subject, purpose, and audience.
8. The pupil can evaluate and edit choice of language in a written expression.
9. The pupil states that mood is the generalized feeling that the setting, tone, and action evoke in the reader.
10. The pupil can write letters to newspapers, businesses, political figures for a variety of purposes.

*Developed by the District Language Curriculum Committee

SCALING PUPIL PROFICIENCIES

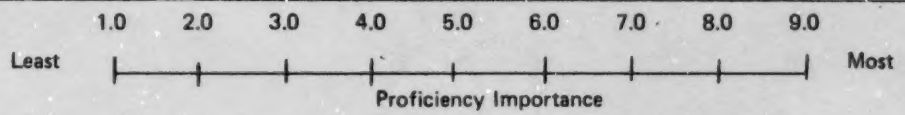
Instructions: Indicate the importance that you place on the identified educational proficiencies by assigning 50 points to them. Because the points are limited, you must decide carefully how you would like to allocate them. For each proficiency, you may commit from one point to a maximum of nine points. Enter your point allocations on the GROUP DATA CHART.

	Point Allocation									
Proficiency	1	2	3	4	5	6	7	8	9	Cumulative Points

GROUP DATA CHART (SAMPLE)

Goal Area _____

	Proficiency	Group _____										Sum	Mean
1	Predictability												
2	Intent												
3	Contracted												
4	Pronunciation												
5	Context												
6	Attitude												
7	Sequence												
8	Evaluate												
9	Mood												
10	Letters												



Review Statement

The first portion of this workshop was concerned with the various functions that should be addressed by program managers to ensure that the instructional program will be responsive to the various school audiences.

Participants, by means of exercises, now have developed a district philosophy, program goals, and objectives. They have classified program objectives by taxonomic level, and have put the objectives into priority order.

The remainder of this workshop will be concerned with the assessment of pupil achievement. Participants will review the various kinds of assessment instruments available and become aware of the advantages and disadvantages of each. They will also spend some time in the development of assessment items, and will critique the developed items. The importance of meeting the needs of pupils requiring alternative instruction in their educational experience is also addressed.

STEP 5: CONSTRUCT/ADOPT ASSESSMENT INSTRUMENTS

Pupil assessment simply indicates whether pupils have learned what they have been taught. The district prescribed courses of study contain objectives which indicate what pupils are to learn. These objectives also provide the basis for pupil assessment. Assessment may be in a generalizable skill, a functional, and/or an applied performance mode, depending on the statement of the objective. (Assessment modes are clarified below.) Particular assessment procedures may consist of academic-related items or situationally-applied items; they may be developed locally or by a test publisher. The degree to which a pupil assessment procedure will be defensible, will, in part, be directly related to how well the assessment correctly and consistently differentiates pupils who have learned what was taught from those who have not. Another important dimension of defending assessment procedure is whether the assessment accurately indicates which pupils can later independently apply what they have learned.

The tests, structured observations, work samples, or other forms of instrumentation used to gether data on pupil performance make up the assessment procedures. The assessment procedures provide the data to make judgments about pupil proficiency. In some circumstances one measure of pupil performance may be adequate verification of a pupil's level of proficiency. In other circumstances multiple measures of pupil performance may be required to verify pupil proficiency.

Constructing instruments to assess pupil proficiency is a difficult task but can be done well by district staffs, from small to large districts, provided they observe principles of test construction. Some districts would prefer to select assessment instruments from among available resources rather than construct their own. Different types of instruments are attractive to districts for varying reasons. Regardless of which type instrument(s) a district might decide to use, it is essential that they select instruments which pertain directly to, and accurately assess, pupil performance in their local courses of study.

Whether districts develop assessment instruments locally or purchase them from another source is not of primary importance. What is critical is that the assessment must relate to the objectives which specify what pupils will be able to do following instruction.

Pupil assessment, as defined in this workshop, emphasizes the individual pupil. **Assessment is the process of forming sound judgments based on data**. This definition of assessment contains terms which should be clarified:

- **Process** is regarded as purposeful activity directed toward an end.
- **Forming** is the process of deliberate, considered action which controls error to the extent it can be foreseen.
- **Sound** implies that judgments are carefully researched and logically defensible.
- **Judgments** are conclusions, deliberately derived from careful examination of data.
- **Data** are summarized information which have been systematically gathered according to stated rules. To the extent that data are free from unsupportable subjective impressions, the validity and reliability of judgments are improved.

Data used to assess pupil proficiencies should be based on procedures designed to relate pupil performance to a prescribed standard. It is not appropriate to compare pupils with one another or to describe their relative proficiencies when the concern is to assure that all pupils attain a minimum standard of proficiency.

There are presently two types of instruments to measure pupil proficiency in such areas as reading comprehension, writing, and computation — norm-referenced and criterion-referenced. The differences between norm and criterion-referenced measures have sometimes been exaggerated to illustrate the inadequacy of norm-referenced measures for individual assessment, and, in some cases, program evaluation. Norm and criterion-referenced measures may differ widely or be identical in all important aspects except construction procedures. Both may have norms, may indicate a performance standard to verify mastery, and may have standardized administration and scoring procedures. The **purpose** for which a measure is constructed, however, is the central variable to consider in selecting appropriate instruments. The purpose for constructing norm and criterion-referenced measures always clearly separates the two types. Norm-referenced measures are developed for the purpose of relating an individual's performance on that instrument to the performance of others; i.e., the norming group. Criterion-referenced measures, conversely, are developed to determine whether an individual can perform acceptably when compared to a preestablished standard.

Criterion and norm-referenced measures can be of the "supply" type (essay, completion) or "selection" type (multiple choice, matching, true-false). The measure that is appropriate is dependent on the nature of the content being assessed, the skill of the person writing the test, and other variables which could influence outcomes or performance on the measure (Wiersma and Jars, 1976). The advantages and disadvantages of different types of measures have been summarized on pages 39 to 40.

RELATIVE ADVANTAGES AND DISADVANTAGES OF ESSAY AND OTHER SUPPLY-TYPE TESTS

Advantages	Disadvantages
1. Easily prepared (Usually has fewer questions)	1. May have relatively low reliability, owing to: a. Limited sampling of learnings b. Subjectivity of scoring
2. Largely eliminates guessing	2. May have relatively low validity, owing to: a. Limited sampling of learnings b. Low reliability
3. Is a direct method of testing some outcomes. Test exercises can closely approximate criterion behavior (i.e., journalism, creative writing)	3. Requires much time of students to write
4. Provides more adequate basis for making inferences about student's level of competency (i.e., ability to define words, write clear explanations of procedures, or complete a geometric proof)	4. Tends to be quickly and carelessly constructed, with the result that: a. Questions are ambiguously stated b. Questions are so general that a student can bluff or "talk around" the subject c. Questions are of unequal difficulty, with no provision for weighting them unevenly in the scoring process d. Selection of questions is not representative of major learnings
5. May give student opportunity to demonstrate his ability to: a. Choose most pertinent and important learnings b. Organize personal knowledge c. Express opinions and attitudes d. Show initiative and originality	5. Tends to be graded without an adequate scoring key, so that results are affected by: a. The "halo effect" (good or bad) of a student's previous level of performance or success on a single question of the test b. Legibility of handwriting c. Errors of spelling and grammar d. Inappropriate emphasis on effectiveness of written expression
6. May be useful for diagnosing incorrect interpretations and partially understood concepts.	
7. May provide insight into instructional ambiguities by revealing frequently misunderstood concepts.	

(adapted from Adams 1964, pp. 330-1)

RELATIVE ADVANTAGES AND DISADVANTAGES OF OBJECTIVE OR SELECTION-TYPE ITEMS

Advantages

1. Makes possible extensive sampling of learnings in a relatively short testing time.
2. Tends to have high reliability as the result of:
   a. Objectivity of scoring
   b. Extensiveness of sampling
3. Is scored objectively, with the result that:
   a. Scoring time is reduced
   b. School personnel are freed from suspicion of partiality
   c. Students' scores are not affected by such extraneous factors as ability to write rapidly and the like
   d. Scoring may be delegated to clerical workers or readers
   e. Statistical analysis of student performance on the test and on specific items is facilitated
   f. At the upper grade levels, tests can be scored by machine and summaries made of student success on each test item
4. Focuses student attention on specific facts or abilities being tested, permitting no evasion.
5. May be valuable for such instructional purposes as:
   a. Pretesting
   b. Diagnostic testing
   c. Individualized self-testing
   d. Testing for application of principles to new situations
6. Adequate test preparation requires thorough awareness of the goals of instruction, major concepts to be developed, facts related to such concepts, and desired level of student competency, by the person(s) constructing the test.

Disadvantages

1. Requires time and skill for adequate preparation.
2. Is of limited value in some subjects (i.e., aspects of language, art, physical education)
3. May stimulate superficial learning of many details because of:
   a. Emphasis placed on recognition of correct answer rather than on remembering
   b. Failure to require the student to organize significant facts and ideas and to reason about them
4. May include ambiguous or misleading questions
5. May result in unduly high scores for intelligent, test-wise students, who have not studied, because of:
   a. Transparent clues in grammar, word form, or phrasing
   b. Insufficient care in the devising of incorrect responses, allowing correct answers to be chosen more by the avoidance of obviously poor choices than by recognition of the correctness of the right response.

(adapted from Adams 1964, pp. 330-1)

Basic Evaluation Instruments

There are four basic kinds of instruments that are available to evaluate learner attainments. They are achievement instruments, questionnaires, observation forms, and logs.

Instrument selection depends on what is to be measured. Each kind of instrument will yield different information.

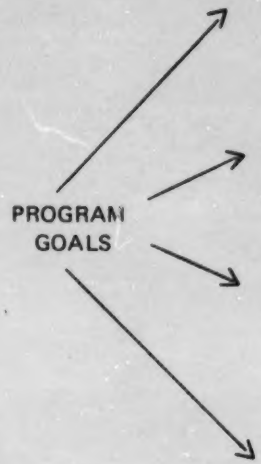
An instrument should be selected for each discrete learning outcome that is expected to result from the program. Skills, abilities, knowledge, and understanding are outcomes best measured by tests. Attitudes, feelings, and appreciations are more appropriately measured by questionnaires and structured interviews. Behaviors, interactions, and practices may be more satisfactorily assessed by means of observation instruments. High-priority objectives will require multiple measures.

Carefull selection and development of instruments for program evaluation help assure that all the information needed to judge how effectively objectives are met will be available when the data-collection effort is complete. (California Evaluation Improvement Project, 1976, p. D-i.)

On the following page is the "Summary of Basic Evaluation Instruments" developed by the California Evaluation Improvement Project. This summary shows the basic types of evaluation measures and their categories, the primary structure of items, and primary kinds of scores they yield. Essay items are also appropriate for assessing pupil achievement; they are not included in the summary but their use is encouraged, particularly in such areas as writing assessments.

A process that is recommended to educators who are or will be assessing pupil proficiency in written expression is: Gottesman, Julia; and Schilling, Mary **A common Ground for Assessing Competencies in Written Expression** Los Angeles: Office of the Los Angeles County Superintendent of Schools, in press.

SUMMARY OF BASIC EVALUATION INSTRUMENTS



Types of Evaluation Measures	Categories of Instruments	Primary Structure of Items	Primary kinds of scores
ACHIEVEMENT	Norm-Referenced Tests Objective-Referenced Tests	. Objective . True/False . Multiple Choice . Matching	Raw Scores Grade Equivalents Percentiles Standard Scores Stanines Percentages
ATTITUDE	Questionnaires	Open-Ended Objective . Yes/No . Multiple Choice . Ratings Mixed	Categories Frequencies Percentages Ratings
INTERACTION	Observations	Open-Ended Objective . Ratings . Present/Absent	Categories Frequencies Percentages Ratings Time
UNOBTRUSIVE	Logs . Referral Reports . Attendance . Cuts . Grades . Diaries	Open-Ended Objective	Frequencies Categories

(California Evaluation Improvement Project, 1975, p. D-81)

Exercise 6: Instrument Selection

CONCEPT: Some districts may conduct programs for which appropriate pupil assessment instruments already exist. In these instances the districts should follow a process that will enable them to select the most appropriate instruments for their particular situation. Resources from which instruments might be selected are: (1) Norm-referenced measures accompanied by a complete set of objectives the test was designed to measure; (2) Criterion-referenced instruments developed by evaluation centers or commercial firms; (3) "Pools" of validated items which are matched to the test specifications prepared by a district.

Selection of norm-referenced measures to be used in a pupil assessment system must be made with care. Selected instruments should always permit direct inferences to be made about pupil performance based on assessment findings. Many norm-referenced measures are not accompanied by information which specifies the particular performance being assessed by particular items. Norm-referenced measures are designed to separate individuals according to their knoweldge or achievement level. Therefore, situations where norm-referenced measures can provide the total basis for pupil assessment would be very difficult to find.

There are some conditions where norm-referenced measures can provide useful information to district personnel. They would be useful, for example, when a district has well explicated courses of study, with all goals and objectives well stated; and

- The selected norm-referenced measure is accompanied by a set of objectives being assessed by the test items; and
- The objectives from the norm-referenced measure match those contained in the course of study; and
- The items on the test assess pupil performance in the way that the course of study objectives indicate that they should be assessed, i.e., generalizable skill, functional or applied performance;
- **Then** the norm-referenced measure could be a useful screening device to identify pupils needing further assessment.

Norm-referenced tests are not designed to be, nor are they appropriate for use as, individual diagnostic instruments.

Criterion-referenced measures are more easily selected for local district application than are norm-referenced measures. Districts with clearly stated instructional objectives and knowledge of what needs to be assesssed in order to verify pupil proficiency can select available criterion-referenced measures which should be readily applicable to their district. Several reputable private firms and government sponsored organizations develop and distribute criterion-referenced measures suitable for district use.

Several organizations commercially produce criterion-referenced measures. Among these organizations are:

1. Instructional Objectives Exchange (IOX), Box 24095, Los Angeles, CA 90024

2. SCORE, Houghton Mifflin Co., 777 California Ave., Palo Alto, CA 94304

3. Comprehensive Achievement Monitoring (CAM), Sequoia Union High School District, 480 James Avenue, Redwood City, CA 94063

4. National Assessment of Educational Progress (NAEP), 300 Lincoln Tower, 1860 Lincoln Avenue, Denver, CO 80203

5. ORBIT, CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 90340

It is possible for districts to send their local objectives and statements of how the objectives should be assessed to test developers and to purchase a collection, or pool, of items that validly assess those objectives. This "item pooling" procedure is attractive to many districts because of the flexibility that it affords in preparing alternate test forms without expanding time and effort spent writing test items. The resulting tests are criterion-referenced.

A complete pupil proficiency assessment system should include multiple measures or a process for verifying professional judgments of pupil performance based on several data sources, if necessary.

Norm and criterion-referenced measures are most commonly thought of as paper and pencil tests. Criterion-referenced measures can also include unobtrusive measures such as observation records, work samples, etc.

PROCEDURE: Abbreviated portions of objectives are listed on the following page. Also listed are selected types of instruments available for pupil assessment. Decide and indicate what type of measuring instrument would be most appropriate to use for each objective cited. Upon completion of this exercise, turn to page 45 and provide the information asked for in the instructions.

SELECTING APPROPRIATE INSTRUMENTS

Instrument Type

A. Norm-Referenced Test

B. Criterion-Referenced Test

C. Questionnaire

D. Observation Record

E. Log

F. Essay

____ 1. Students participating in the Career Education program will have positive attitudes about the program...

____ 2. Twenty-five percent of the students will achieve one standard deviation above the national mean in math...

____ 3. Students will interact in a positive social manner during class activities.

____ 4. Given a list of South American countries, students will be able to list the capital of each country.

____ 5. The number of discipline referrals to the principal will be reduced by fifty percent...

____ 6. Students will check out books in category "A" more frequently than books in category "B"...

____ 7. Students will apply the rule that the speed with which an object changes position is the distance moved per unit of time.

____ 8. Students will perceive and describe visual forms in terms of structure and expression.

____ 9. Teacher effectiveness in task analyzing instructional objectives will increase...

____ 10. The majority of parents with students attending school "X" will be aware of the auxiliary services available through the school...

APPROPRIATE FOR PUPIL ASSESSMENT

Which of the objectives on page 45 could be considered for inclusion in assessing learner proficiency for purposes of the Hart Law or AB 65?

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

COMPARISON SHEET
SELECTING APPROPRIATE INSTRUMENTS

1. C
2. A
3. D
4. B
5. E
6. E
7. B
8. B
9. D
10. C

APPROPRIATE FOR PUPIL ASSESSMENT

1. no
2. no
3. no
4. no
5. no
6. no
7. yes
8. yes
9. no
10. no

Exercise 7: Developing Instruments Specifications and Items, Assessing Item Adequacy

CONCEPT: Some districts will choose to develop assessment instruments which are directly related to their local curriculum. They may choose to develop their own instruments for such reasons as: measures particularly suited to their curriculum may not be available, and selection from among available measures still requires local validation if those measures are to be used to verify attainment of proficiency standards.

Before actual assessment items are written, it is necessary to have a good "blueprint" or set of instrument specifications. Without these specifications, each person developing an item to measure a certain objective might construct an item very different from those of other developers.

> A basic premise of pupil proficiency assessment is that pupils who meet prescribed performance standards have highly similar skills and knowledge. Therefore, it is important for the instruments to have homogeneous items. Homogeneity of items is proportionate to the clarity of the parameters stated in the test specifications.

It is important to note that if districts decide to select instruments from among those available, they must still prepare test specifications to guide selection, which is the most difficult part of instrument development. Therefore, many districts may elect to develop their own tests and supplement them as the opportunities become apparent, rather than to initiate a "search and validate" process.

Development of pupil proficiency assessment instruments is a three-stage process which includes planning, constructing, and verifying.

Planning: There is always a need for proper planning when building assessment instruments. When completed instruments have been carefully conceptualized before work begins, progress will be more rapid and there will be a much lower risk of producing useless tests. The staff should examine the needs for, and purposes of, the instruments, the specific content to be assessed, and how they will be presented or formatted. The political aspects of how the instruments will be used and who will be affected by their use needs to be considered. Resources available to assist in developing the instruments may require that priorities be determined regarding what will be assessed. A broad base of input from the various school audiences should be received when deciding on the content to be assessed. For purposes of verifying that pupils have mastered the basic skills, it is not practical to assess pupil performance on virtually every objective in the courses of study. Only those objectives which represent critical learning that should be required of all students should be assessed.

For more general monitoring of pupil progress, the curricula should be examined to identify those broader objectives which subsume several previously learned skills or concepts. Assessment would be done at those broader "break points" in the curricula rather than involving pupils and staff in assessment of minutiae.

Test format and item type should be considered early because of their implications for scoring procedures and interpretation of findings. Most medium to smaller sized districts will be restricted somewhat by the level of funding available for optical scanning devices, computer scoring, or other mechanical applications to testing. These smaller districts should be careful to limit their assessment procedures to a manageable undertaking and not overextend their resources.

Constructing: constructing assessment instruments requires early identification of instrument specifications. These specifications set the limits within which those writing the instrument items must stay. When assessing pupil proficiency at a particular level, it is imperative that items on any instrument be homogeneous in scope and difficulty. If the items are homogeneous, logical inferences can be made about the level of proficiency attained by a pupil that is reflected in the scope. If the items are widely varied, it is difficult or impossible to determine what effect guessing on the test, or other chance influences, might be reflected in the score. Some test construction experts also recommend that items should be of like form; that is, all items should be of the generalizable skill, functional, or applied performance testing type. When items are also homogeneous in form, score interpretations are more easily made. Since the concept of assessment includes the notion that professional judgment is involved in data interpretation, homogeneity of form is not imperative. It seems that districts could defend a judgment that a pupil is or is not proficient based on input from a variety of data sources, provided the data were objectively obtained and relate to prescribed standards. Instrument specifications consist of the following:

1. A general description of what and how a behavior is to be assessed. These are sometimes referred to as assessment context and mode descriptors.
2. An example of an assessment item.
3. A set of statements regulating how an item may appear. These statements are referred to as "stimulus attributes" because they are characteristics of the item (stimulus) to which the pupils taking the test or completing tasks on an observation schedule, etc., will respond. Stimulus attributes define the parameters of modes, means, and materials that can be included in the **items**.
4. A set of statements defining which responses to an item are, or are not, satisfactory. These statements are referred to as "response attributes" because they are characteristics of correct solutions or answers to problems or questions that differentiate between the responses that **are** awarded credit from those that are **not** awarded credit. Response attributes define the parameters of modes, means, and materials that can be included in the **responses**.

When building instrument specifications, first build an item that is typical of those desired, and proceed to the development of stimulus and the response attributes. As assessment items are developed, attend to the details of scoring procedures and, perhaps, classify the item according to the level of mental, affective, or physical functioning required by the item.

Before generating assessment items, it is advisable to circulate the test specifications among those developing instruments to detect possible errors and to verify the soundness of the written specifications. Those writing the items should agree that the specifications for which items will be written are sound.

Principles of item development related to particular assessment procedures, item type, format, and standards of item development must be observed.

Key rules for item construction which have practical application are:

A. The Stem:
   1. If the stem is a question, start it with an interrogative word.
   2. Don't force the stem into the form of a direct question if an incomplete statement is more appropriate.
   3. Use clear and simple language.
   4. If the item is measuring vocabulary, the highest level of language used in the item should be below that considered appropriate for the grade being tested.
   5. Avoid negatives or double negatives. If used, capitalize all letters in the negative word.
   6. Each stem must have only one central problem. The examinee should be able to determine the answer without looking at the responses.

B. The "Answer":
   1. There should be only one correct answer to an item.
   2. Where examinees are to pick the "best" answer, there must be a "best" **upon which experts would agree.**
   3. Do not express the correct answer more carefully or at greater length than the alternatives.
   4. Avoid putting key words from the stem in the correct answer only.
   5. Make certain the correct answer is no more consistent grammatically or semantically with the stem than are any of the alternatives.

C. The Alternatives or Distracters:
   1. All items need not have the same number of alternatives unless "formula" scoring is to be accomplished.
   2. Alternatives should be grammatically consistent with the stem.
   3. Alternatives should be grammatically parallel.
   4. Distracters should be plausible and attractive to the examinees who does not know the right answer.
   5. Choices should be arranged in logical order unless the order reveals the answer.
   6. Numerical responses should be from smallest to largest number, or the reverse.
   7. Single-word responses should be alphabetized unless there is logic for another order, such as months of the year.
   8. Lengthy responses should be arranged in order of their length.
   9. Choices that are identical with names of things on a graph should be ordered as they are on the graph.
   10. Distracters should be independent and mutually exclusive.
   11. Symbols used to identify alternatives should be such that they cannot be confused with the content of the responses.
   12. If choices are letters, identify the alternatives with numerals; if responses are numerals, use letters. Do not repeat the same letters or numerals used to identify answer responses.
   13. Do not force a fourth or fifth choice into an item which logically can have only three choices.

D. Keying the Item:
   1. There should be about the same number of A, B, C, and D keyed responses in a four-choice test.
   2. The keyed response of an item should not consistently be the longest response or the shortest response.
   3. The same keyed response should not appear in a long series of consecutive items.

E. Miscellaneous
   1. Avoid words and idioms whose connotations vary with locale.
   2. Always keep in mind what the test is meant to measure.
   3. Test items should be numbered consecutively if possible.
   4. Check all facts stated in the item, using reliable reference sources.
   5. Arrange items, when possible, in order of difficulty from easy to hard. When items occur in groups, arrange groups in order of difficulty.
   6. Avoid items that are based on rote verbal association. (i.e., Q: What is the sun? A: A large celestial mass of igneous gas. Correct response is taken right from the text).

F. Directions:

The directions for administering and taking a test afford all those being assessed a valid opportunity to do as well as they are able. Perhaps it is sufficient to say that the directions should not affect the outcome of the measurement. One taking the test should not be penalized or assisted because of the directions. It is helpful to pattern directions after an available test which has been found to be clear and useful.

After the measures have been constructed, they should be field tested and edited to final form.

Verifying: Assessment instruments must be appropriate and accurate. Because of the seriousness of consequences for pupils who do poorly on the districts' assessment instruments, it is essential that their tests be highly defensible. Therefore, districts will want to use instruments which accurately and consistently measure what they are intended to measure.

To verify instrument acceptability, each item should be inspected to assure that it contributes to differentiating pupils who are proficient on the skills or knowledge being measured from those who are not. It is important to field test the instruments in situations where they will later be used.

The procedure generally followed when field testing items is:

- Assemble constructed test items.
- Format the items into an instrument that is proposed for use in the assessment system.
- Administer the field test instrument.
- Compile the data from the field test.
- Analyze the date, refine/reject items.
- Include verified items in the assessment system.

To assist in verifying locally-developed instruments, it is sometimes helpful to probe the items by asking questions about them. The following probes have been organized according to item type and mechanical features of a test.

A. Essay Questions
   1. Are essay questions restricted to measurement of objectives not readily measured by other item types?
   2. Are essay questions framed around specific problems, adequately delimited in scope?
   3. In general, is use made of a large number of brief essays rather than one or two extended essays?
   4. Does each essay question indicate clearly and accurately the desired extent and depth of the answer?
      a. Does the question indicate clearly how many reasons, examples, arguments, etc. are expected for full credit?
      .b. Do the directions to "explain," "outline," "state," "compare," etc., indicate accurately the type of answer that will receive full credit?

B. Completion Items
   1. Is each statement sufficiently unambiguous to limit the correct answer to one or two specific words or phrases?
   2. Is an excessive number of blank spaces avoided?
   3. Does the omitted part of each question come at the end or near the end of the statement?
   4. Are the items free from extraneous clues due to grammatical structure, length of blank space, etc.?
   .5. Is use of the completion form generally avoided when items are essentially true-false in nature?
   6. Do computational problems indicate the expected degree of precision in the answer?
      a. Does the question indicate clearly the extent to which approximations or fractional answers are to be rounded?
      b. Does the question indicate clearly whether units (such as square yard or feet per second) are to be included by the pupil in his answer?

C. True-False Items
   1. Is each item definite and unambiguous in meaning?
   2. Are the items based upon statements that are absolutely true or false, without qualifications or exceptions?
   3. Has the central point of each question been highlighted by placing it in a prominent position?
   4. Are the items free from statements that are partly true and partly false?
   5. Are the items expressed as simply and directly as possible, without many qualifying clauses?
   6. Are trick questions avoided?
   7. Are negative questions avoided so far as possible?
   8. Are the items free from specific determiners such as "always" and "usually?"
   9. In the modified true-false type of item, is the word to be corrected clearly indicated by underlining or special type?

D. Multiple Choice Items

1. Is use of the direct question or imcomplete statement form consistent with the most effective presentation of the individual items?
2. Are the items presented in clear and simple language, with vocabulary kept as simple as possible?
3. Does each item have one and only one correct answer?
4. Is each item concerned with a single central problem?
5. Is the central problem of each item stated clearly and completely in the stem?
6. Does the stem, so far as possible, include all words repeated in the responses?
7. Are negative statements avoided?
8. Do the responses or choices come at the ends of the incomplete statements?
9. Are the responses grammatically consistent with the stem and parallel with one another in form?
10. Are all responses plausible and attractive to pupils who lack the information or ability tested by the item?
11. Are the responses, so far as possible, arranged in numerical or logical order?
12. Are the responses independent and mutually exclusive?
13. Are the items free from extraneous clues due to grammatical inconsistencies, rote verbal associations, length of response, etc.?
14. Is the "none-of-these" option used only when appropriate?

E. Matching Items

1. Do the premises and responses constitute homogeneous lists, each grouped around a single concept?
2. Are the lists of premises and responses relatively short?
3. Are the matching lists conveniently arranged?
   a. Are the longer, more complex statements used as premises and the shorter statements as responses?
   b. Wherever possible, are the responses arranged in some logical order to simplify matching?
4. Do the directions indicate clearly the basis upon which the lists are to be matched?
   a. If a single premise is to be matched with several responses, is the pupil advised that this is permissible?
   b. If a single response is to be matched with several premises, is the pupil advised that this is permissible?
5. Are the matching lists free from extraneous clues due to grammatical construction, rote verbal associations, etc.?
6. Is the list of responses longer than the list of premises to preclude guessing by elimination?

Mechanical Features of the Test

A. Item Format
   1. Are the items in the test numbered?
   2. Is each item complete on a page?
   3. Does the reference material for an item appear on the same page as the item or on a facing page?
   4. Are the item responses arranged to achieve both legibility and economy of space?

B. Scoring Arrangements
   1. Has consideration been given to the practicability of a separate answer sheet?
   2. Are answers to be indicated by symbols rather than by underlining or copying?
   3. Are answer spaces placed in a vertical column for easy scoring?
   4. If answer spaces are placed at the right of the page, is each answer space cearly associated with its corresponding item?
   5. Are the answer symbols to be used by the pupils free from possible ambiguity due to careless penmanship or deliberate hedging?
   6. Are the answer symbols to be used by the pupils free from confusion with the substance or content of the responeses?

C. Distribution of Correct Responses
   1. Are correct answers distributed so that the same answer does not appear in a long series of consecutive questions?
   2. Are correct answers distributed to avoid an excessive proportion of items in the test with the same answer?
   3. Is patterning of answers in a fixed, repeating sequence avoided?

D. Grouping and Arrangement of Items
   1. Are items of the same type requiring the same directions grouped together in the test?
   2. Where juxtaposition of items of markedly dissimilar content is likely to cause confusion, are items grouped by content within each item type grouping?
   3. Are items generally arranged from easy to more difficult within the test as a whole and within each major subdivision of the test?

E. Designating Credit Allowances
   1. Are credits indicated for the major sections of the test?
   2. Is the credit allowance for each item clear to the pupil?
   3. Where questions have subdivisions, especially in essay questions, are credits indicated for each of the parts of the question?

F. Directions for Answering Questions
   1. Are simple, clear and specific directions given for each different item type in the test?
   2. Are directions clearly set off from the rest of the test by appropriate spacing or type style?
   3. Is effective use made of sample questions and answers to help clarify directions for unusual item types?

G. Correction for Guessing
   1. If deductions are to be made for wrong answers, are pupils instructed not to guess?
   2. If no deductions are to be made for wrong answers, are pupils advised to answer every question according to their best judgment?

H. Allowing Choice of Items
   1. Is the degree of choice sufficiently limited, and are the questions from which choice is allowed sufficiently similar in difficulty, to maintain reasonable comparability of pupils' scores?
   2. Are the directions covering choice prominent, clear, and explicit?
   3. Is choice exercised within relatively small groups of items rather than among items?

I. Printing and Duplicating
   1. Has the test been duplicated to provide individual pupil copies?
   2. Is the test free from annoying and confusing typographical errors?

J. Is the legibility of the test satisfactory from the viewpoint of type size, adequacy of spacing and clarity of printing?
   3. Is the legibility of the test satisfactory from the viewpoint of type size, adequacy of spacing and clarity of printing?
   4. Is the length of line neither too long nor too short for easy comprehension?

Exercise 7; Part A; Instrument Specifications and Item Development

PROCEDURE:

1. Participants review pages 48 through 55.

2. Participants review the sample instrument specifications shown on the following 3 pages.

3. Discuss the instrument specifications, recognizing that the first set (pages 54 to 55) is written for an applied skill and the second set (page 56) is written for a generalizable skill type of performance objective.

4. Working in teams of 4-6 participants per group, each team develops **one** set of instrument specifications and records them in the space provided (page 57).

5. After all groups have developed a set of instrument specifications and recorded them on page 57, the groups exchang specifications. The receiving group develops three items appropriate for that set of specifications and records them in the space provided (page 58).

6. Each group returns the specifications, together with their developed items, to the group that developed the specifications.

7. The group that developed the specifications reviews the items and judges the degree to which they match their specifications and records their evaluation in the space provided (page 58).

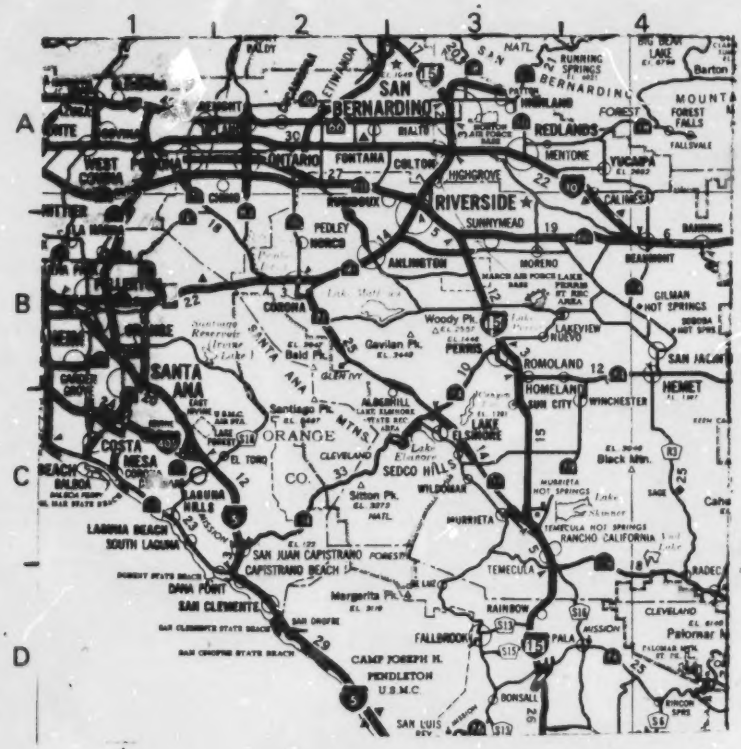
FUNCTIONAL SKILL

Objective: Upon exit from high school, students will be able to interpret a map when locating points of origin and destination, and determine the distance between those points.

General Description: When presented with a mercator projection map utilizing a uniform scale which contains signs, graphs, charts, and diagrams, the respondent will be able to locate preestablished starting and ending points and calculate the distance between those points to the nearest mile.

Sample Item

Directions: Beginning at Riverside (A-3), what is the shortest distance to San Juan Capistrano (C-2)?


1
2
3
4
A
B
C
D
SAN BERNARDINO
RIVERSIDE
REDLANDS
ONTARIO
FONTANA
COLTON
HIGHGROVE
MENTONE
YUCAIPA
CALIMESA
SUNNYMEAD
MORENO
BEAUMONT
BANNING
PEDLEY
NORCO
ARLINGTON
CORONA
SANTA ANA
ORANGE
CO.
COSTA MESA
NEWPORT BEACH
LAGUNA BEACH
SOUTH LAGUNA
LAGUNA HILLS
SAN JUAN CAPISTRANO
CAPISTRANO BEACH
DANA POINT
SAN CLEMENTE
TEMECULA
MURRIETA
WILDOMAR
SEDCO HILLS
LAKE ELSINORE
PERRIS
ROMOLAND
HOMELAND
SUN CITY
WINCHESTER
HEMET
SAN JACINTO
GILMAN HOT SPRINGS
LAKEVIEW
NUEVO
GAVILAN PK.
WOODY PK.
BALD PK.
SANTIAGO PK.
ALBERHILL
LAKE ELSINORE
MURRIETA HOT SPRINGS
TEMECULA HOT SPRINGS
RANCHO CALIFORNIA
BLACK MTN.
SAGE
RADEC
RAINBOW
FALLBROOK
PALA
MISSION
CLEVELAND
PALOMAR MTN.
CAMP JOSEPH H. PENDLETON U.S.M.C.
SAN LUIS REY
BONSALL
MARGARITA PK.
SUTTON PK.
MTNS.
BARON
MOUNTAIN FOREST FALLS
FALLSVALE
BIG BEAR LAKE
RUNNING SPRINGS
FOREST
HIGHLAND
SAN BERNARDINO NATL. FOREST
MARCH AIR FORCE BASE
PERRIS LAKE
PERRIS ST. REC. AREA
SANTA ANA RIVER
SANTA ANA CANYON
LAKE MATHEWS
LAKE FORBES
LAKE MORENA
LAKE HODGES
SAN DIEGUITO RESERVOIR
SAN ONOFRE STATE BEACH
DOHENY STATE BEACH
SAN CLEMENTE STATE BEACH
SAN LUIS REY
SAN DIEGO
SAN DIEGO CO.
SAN DIEGO NATL.
SAN BERNARDINO NATL.
SAN ONOFRE
SAN LUIS REY
SAN DIEGUITO

Stimulus Attributes:

1. The map must not exceed 8½ x 11 inches.

2. The scale of miles per inch used on the map is identified.

3. The map will clearly show North, and other directions will be true to scale from that indicator.

4. The map will show miles between points on the map using consistent markings, and will be supplemented by a scale of miles which would enable one to determine distance between points.

5. Town, city, street, highways, and other predominant geographical locations will be accurately identified on the map.

Response Attributes:

1. Four answer options will be presented, one of which corresponds to the shortest route and distance shown on the map. All routes will follow highways which originate at the starting and ending points, but the mileage calculations may be in error.

2. Incorrect answers may be:

   a. The route and mileage that would result from selecting a longer presented route.

   b. The distance selected on the basis of the smallest mileage figure, but without validating the answer.

   c. A route which visually may appear to be a shorter route.

   d. An obviously longer route associated with a short mileage figure.

3. The order of correct and incorrect answers will be random.

GENERALIZABLE SKILL

Objective: When presented with horizontally or vertically aligned multiplication problems and an array of answers, one of which is correct, the respondent will be able to select the correct answer from an array of possible choices.

Sample Item:

Directions: For each problem, mark an "X" next to the correct answer to the multiplication problem.

$$\begin{array}{r} 38 \\ \times 47 \\ \hline \end{array}$$

___A 1436

___B 1705

___C 1786

___D 1796

Answer C

Taxonomic Level: Knowledge

Stimulus Attributes:

1. The multiplication problems may be horizontally or vertically aligned.
2. Two-digit, base ten, whole numbers are used in the multiplier and multiplicand.
3. Only one correct answer will be presented.
4. Answer choices will be to the right of the problem and vertically aligned.
5. All problems will include regrouping.

Response Attributes:

1. Four answer options will be presented, one of which is accurate.
2. Incorrect answers may be:
   a. The answer choice shows multiplication completed without regrouping.
   b. The answer choice shows multiplication completed with number transpositions so the wrong digits are regrouped as tens.
   c. The answer choice shows errors in addition.
3. The correct and incorrect answer options are arranged in ascending numerical order.

Now each team should develop test specifications and items for an applied or generalizable skill of interest to that group. **Develop only one set of specifications.**

Instrument Specifications

Items:

a.

b.

c.

Evaluation of Item Adequacy:

Item a:

Item b:

Item c:

Exercise 7; Part B: Assessing Item Adequacy

PROCEDURE:

1. Participants, working independently, review the test items found on page 60.

2. Determine whether each item is adequately constructed in using criteria stated on pages 50 and 51.

3. Mark an "X" for each adequate item in the space provided.

4. After completing the exercise, participants should compare their answers with the Comparison Sheet found on page 63.

PROCEDURE: Indicate whether the following items are adequate or inadequate. Mark an "X" beside the adequate test items.

____1. Jack had two pitchers filled with orange juice. He poured all of the orange juice into one larger pitcher. If one of the smaller pitchers held 2 liters of juice and the other smaller pitcher held 4 liters, how much orange juice did Jack pour into the larger pitcher?

Answer: _____

____2. Mark the sentence that transforms the underlined sentence from passive voice to active voice.

The lawn was mowed by Ernie so that Dan could rest.

____(a) Dan mowed the lawn for Ernie so that he could rest.

____(b) Dan was rested while Ernie mowed his lawn.

____(c) Ernie mowed the lawn so that Dan could rest.

____3. Determine whether the statement is true or false.

T F "Common law" translates reason into justice and protects all citizens equally.

(Circle one)

____4. Select the correct answer.

The President may extend diplomatic recognition to:

____(a) a country only if the Secretary of State approves.

____(b) a foreign diplomatic representative after securing approval from the Senate Foreign Relations Committee.

____(c) both of the above

____(d) neither of the above

____5. Complete the sentence.

_____ and _____ found the passage in their explorations of the northwestern territory.

Exercise 7: Comparison Sheet — Assessing Item Adequacy

1. Inadequate. "Other-supply" type, completion item; directions missing.

2. Adequate. "Selection" type, multiple choice item; (could have more distractors).

3. Inadequate. "Selection" type — true-false items; compounded item; too inclusive, i.e., "all".

4. Inadequate "Selection" type — multiple choice items; distractors are not plausible.

5. Inadequate. "Other-supply" type completion item; excessive number of blanks. Key term identifies one answer. Generally poor.

Setting Performance Standards: Meeting the Needs of Pupils With Diagnosed Learning Disabilities

Recent legislation requires local districts to provide guidance and educational assistance to students who are not able to function acceptably within the context of the regular school program. This requirement is consistent with the California Master Plan for Special Education and Public Law 94-142, the Education for all Handicapped Children Act of 1975. It appears that districts able to comply with P.L. 94-142 would have no difficulty meeting the requirements of AB 65 or AB 3408. These requirements are:

A) A statement of the present levels of educational performance of each child.

B) A statement of annual goals, including shorter instructional objectives.

C) A statement of the specific educational services to be provided to identified students, and the extent to which each student will be able to participate in regular educational programs.

D) The projected date for initiation and anticipated duration of these services.

E) Appropriate objective criteria and evaluation procedures and schedules for determining, on at least an annual basis, whether instructional objectives are being achieved.

STEP 6: SET PERFORMANCE STANDARDS

Performance standards provide an objective indicator of the level of skill or knowledge attained by a pupil. Ideally, a district's assessment system will have a range of performance standards identifying levels of pupil progress from entry into (kindergarten or grade 1) to exit from school (grade 12). Such a system would permit a staff to verify pupils as proficient at relative levels that could be interpreted in terms of the district's courses of study. It should be noted that an assessment system which encourages comparisons among pupils, rather than to determine the proficiency level of particular pupils, has little value.

How to set performance standards is a matter that should be decided after considering several important factors such as:

- Standard setting is ultimately done through use of human judgment and, therefore, cannot avoid being subjective.
- Standards should be based on data so that they will not be regarded as arbitrary and/or capricious.
- Various school audiences (parents, pupils, community members, etc.) may have different aspirations, values, or expectations related to education and what proficiencies are necessary for pupils to attain.
- Misclassification of pupils as proficient when they are not, or as not proficient when they are, is a problem that districts should attempt to control as much as possible.
- The relationship between standards of excellence and minimum acceptable performance is most easily explained in the context of an assessment system which accounts for the full range of proficiencies a district is attempting to teach.

Present law places emphasis on setting minimum standards of performance at grade eight and for high school graduation. The standards that apply to grade eight pupils are intended to be an early warning to pupils that they may not later qualify for graduation unless they improve their performance level(s). Hopefully, if pupils have been involved in deciding what they need to learn and have advance knowledge of the performance level expected of them, they will be better motivated to achieve the prescribed standards. Parents also would know specifically what their children must learn and can demonstrate a more positive value for those learnings.

Exercise 8: Determining Performance Sandards

CONCEPT: Before pupils can be verified as proficient, it is necessary to determine what performance standards they must attain.

Performance standards are used to describe a relative level of knowledge or proficiency. For example, a young child reads classroom lessons well but is not able to read an advanced philosophical treatise; that child would be considered to be proficient at that grade placement, but not at the college level.

Since the performance standard may very depending on circumstances, it is essential that the standard(s) be set before any judgments about pupil proficiency are made.

Techniques used by districts to arrive at standards of pupil performance depend on the kind, quality, and amount of data available to districts. Three methods that districts might consider are:

1. **Reasonable Expectancy:** When districts have no data to indicate how well pupils should perform on their assessment instruments, they must rely on estimated standards. Although mastery implies 100 percent attainment, it is unreasonable to expect that the majority of pupils would be able to attain this level of mastery. It is necessary, therefore, to determine what level of performance should be expected and make that level known.

2. **Consensus:** When needs assessment data are available, a team of experts can set performance standards for the various pupil populations.

   This team of experts could include representatives from such groups as curriculum specialists, teachers, evaluators, pupil personnel staff, and administrators.

3. **Refinement of Performance Standards From a Baseline:** When districts have valid field test or baseline data, they can set performance standards that are consistent with actual pupil performance on the districts' assessment instruments.

PROCEDURE: This exercise is presented in two parts. For each part participants will use separate data displays. Following are two exhibits of possible computer printouts showing two approaches to finding the proportion of students who successfully mastered items on a test.

On Exhibit No. 1, the same items were presented to students ranging from sixth through eleventh grades. Ninety percent is the desirable mastery level for each item. Study the data and decide at which grade level the item would be appropriately placed.

On Exhibit No. 2, it is assumed that a curriculum committee prepared an end-of-the year computational skills test for grade levels eight and eleven. The printout reflects data from a test administered to eleventh grade students.

Exercise 8, Part A:

1. Participants review Exhibit No. 1

2. Participants locate at which grade level the data indicates the items should be placed.

3. Participants record in column (7) whether the item was appropriately placed; and, if not, at what level it should have been placed.

4. Participants compare their responses with the comparison sheet (page 66) and discuss discrepancies with other participants.

EXHIBIT NO. 1

Goal Area: Addition-Subtraction of Rational Numbers

(1) Skill/Knowledge	(2) Competencies	(3) Question	(4) Proportion of Responses Per Choice	(5) Grade	(6) Grade Placed	(7) Appropriately Placed Yes/No If no, at what level
Determine the sum of two or more rational numbers in fraction form	Involving two numbers less than 1 to sum less than or equal to 1, or possible reducing sum.	Name the sum $\frac{3}{7} + \frac{2}{7} =$ or $\frac{3}{7}$ $+\frac{2}{7}$	(A) (B) (C) (D) $\frac{5}{28}$ $\frac{5}{7}$ $\frac{6}{7}$ $\frac{6}{28}$ .11 .65 .20 .04 .07 .75 .14 .04 .10 .85 .04 .01 .05 .91 .03 .01 .04 .93 .03 .0 .01 .98 .01 .0	6 7 8 9 10 11	7	
Determine the difference of two rational numbers in fraction form.	Involving a mixed fraction sum and the addend less than 1.	Name the difference $3\frac{2}{5} - 1\frac{1}{2}$ or $3\frac{2}{5}$ $-\ 1\frac{1}{5}$	(A) (B) (C) (D) $1\frac{9}{10}$ $2\frac{1}{10}$ $1\frac{1}{5}$ $2\frac{1}{5}$ .40 .38 .10 .12 .55 .31 .04 .10 .70 .20 .03 .07 .80 .15 .02 .03 .92 .07 .01 .00 .97 .02 .0 .01	6 7 8 9 10 11	8	

Exercise 8,Part A: Comparison Sheet Determining Performance Standards

No

Grade 9

or examine possible appropriate curriculum revisions

No

Grade 10

or examine possible appropriate curriculum revisions

Exercise 8, Part B:

1. Participants review Exhibit No. 2

2. Participants identify those items that the data indicate would be homogeneously grouped above the criterion level (90%).

3. Participants record their findings in the right hand column.

4. Participants compare their responses with the comparison sheet (page 69) and discuss discrepancies with other participants.

EXHIBIT NO. 2

Grade Level Competency	Test *Item No.	Proportion of Students Who Accurately Answered the Item	Check ( ) Those Items That Would Be Homogeneously Grouped Above The Criterion Level (90%)
The Student:			
knows the symbols	1	.95	1 _____
for mathematical	2	.93	2 _____
operations	3	.97	3 _____
	4	.91	4 _____
	5	.98	5 _____
is able to add	6	.83	6 _____
whole numbers involving	7	.92	7 _____
3-digit numerals	8	.76	8 _____
with renaming	9	.93	9 _____
	10	.95	10 _____
is able to subtract	11	.84	11 _____
whole numbers	12	.77	12 _____
involving 4-digit	13	.91	13 _____
numbers with	14	.43	14 _____
renaming	15	1.00	15 _____
is able to	16	.65	16 _____
multiply whole	17	.92	17 _____
numbers	18	.85	18 _____
	19	.98	19 _____
	20	.90	20 _____
is able to solve	21	.35	21 _____
simple	22	.60	22 _____
two-step	23	.92	23 _____
problems	24	.38	24 _____
encountered in	25	.94	25 _____
everyday life			

*Having five items per competency is for illustrative purposes only and does not suggest an optimal number of items for a criterion measure of a competency.

Exercise 8; Part B: **Comparison Sheet Determining Performance Standards**

1	.95
2	.93
3	.97
4	.91
5	.98
6	.92
7	
8	.93
9	.95
10	
11	
12	
13	.91
14	
15	1.00
16	.92
17	
18	.98
19	.90
20	
21	
22	.92
23	
24	.94
25	

STEP 7: DEVELOPING MONITORING AND REPORTING PROCEDURES

The pupil proficiency assessment system will be a pervasive part of the districts' operation. It will produce data that will be of interest, at some level of specificity, to all the school audiences. For example:

- Pupils will need to know how well they are progressing in the studies, including progress in the basic skills required for graduation.
- Parents will need to know how their children are progressing academically.
- School site staff will need to know which programs are having greatest success, with which pupils.
- Curriculum specialists, with assistance from other school staff, will need to revise/refine curriculum based on assessment data.
- The central offfice will need to know whether existing administrative regulations and procedures result in development of viable curriculum development and implementation.
- The governing board will need to know which policies are being successfully implemented, which need to be reexamined for possible revision or deletion, and which areas may require new policies.
- The community will want assurances that pupils have acquired proficiencies in skills required by business, industry, or training or education centers.

How districts record pupil proficiency attainments and report them to the various audiences is a key issue in proper management of their assessment systems.

If districts have conceptualized how they will manage their proficiency assessment systems, and how they will collect and report assessment data, the forms that will facilitate these tasks will be developed more easily.

Following are examples of a district management plan and reporting forms that might assist districts in examining their monitoring and reporting procedures.

Exhibit No. 1 shows the overall plan for managing assessment information as it relates to pupil proficiency. The plan suggests the kinds of information that the various school audiences might be provided to facilitate communication and decision making.

Exhibit No. 2 is designed to provide pupil proficiency data for the total district. Interested groups can easily determine the progress being made by pupils in each school.

Exhibit No. 3 is a system for summarizing pupil progress in individual schools. The system is easy to maintain, requiring only that the teachers check each objective met by individual pupils and recording the date that the student has been certified as proficient in the particular goal area.

Exhibit No. 4 is a form designed to verify a pupil's eligibility to graduate. In addition to verifying pupil proficiencies in reading, language, and computation, the form is designed to show pupil proficiency in other course areas required by the school district.

Exhibit No. 5 is a form designed to document conference follow-up activities.

1. The pupil who does not meet the district prescribed standards in basic skills.

2. A report of student performance on proficiency assessment.

3. How the pupil and the pupil's parent/guardian were notified. Includes school staff effort(s) to contact parent/guardian by telephone to reinforce the written communication.

4. The responsible school official who arranged the **required** conference.

5. Who attended the conference (signatures probably provide the best verification of attendance) and when the conference was held.

6. What educational action plan was discussed to assist the pupil attain the district prescribed proficiencies.

7. The school official responsible for instructional follow-up

8. Date supplementary instructional activities began and relevant comments on pupil performance.

EXHIBIT NO. 1

ASSESSMENT INFORMATION MANAGEMENT PLAN

Category of Information	Required Information	Communication Mechanism	Reported to *(specify e.g. - Pupils - Parents - Staff - Administration & Governing Board - Others*
Administrative Regulations	Legal Base and Governing Board Policy Regarding the Assessment System	Pupil Handbook Staff Handbook	
Organization of Assessment System	Administration Calendar Make Up Schedule Special Assessment Arrangements Subject Matter Areas Covered Proficiencies Required Minimum Standards Findings; Individual and Aggregate	Pupil Handbook Staff Handbook Press Release Board Policies Administrative Regulations Others	
Instructional and counseling implications	Matrix showing proficiencies covered in each department by course and grade/level offered Alternate modes of instruction Verification of pupil proficiency	Graduation Eligibility Verification School Data Summary Pupil Monitoring records	
Proficiency conference	Record of required pupil conferences	Assessment follow-up record	

81

Exhibit No. 2

DISTRICT PUPIL PROFICIENCY DATA SUMMARY

District_____ Date_____

Goal Area: Computation

School	N.	Total No. Prof.	% Prof.	Per Cent Proficient on Each Objective 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A																				
B																				
C																				
D																				
Total																				

Exhibit No. 3

SCHOOL PUPIL PROFICIENCY DATA SUMMARY

School: _____ Date: _____

District: _____

Student	Objective																	Proficiency Verification Date
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Adams, J																		
Barnfield, N.																		
Cameron, A.																		
Dill, B.																		
Franks, M.																		
George, C.																		
Gregory, A.																		
Matthews, Y.																		
Mornan, N.																		
Rodriguez, A.																		

No. Students Proficient = _____

Percent Students Proficient = _____

Exhibit No. 4

GRADUATION ELIGIBILITY VERIFICATION

Pupil _____ Date: _____

School _____

Principal _____

Minimum Proficiency Areas	Score	Verification Date

Course Requirements	Units/Score/Grade	Completion/Verification Date

Exhibit No. 5

ASSESSMENT FOLLOW-UP RECORD

Assessment Finding		Conference Notification			Assessment Conference				Instructional Follow-up	
Pupil Name	*Assessment Findings Date	(Guardian) Parent/Pupil Notification: Written	(Guardian) Parent/Pupil Notification: Telephone	Responsible School Official	Parent (Guardian) / Pupil/School Staff *Conference: Participants	Parent (Guardian) / Pupil/School Staff *Conference: Date	Action Taken	Person Responsible	Date Begun	Comments

STEP 8: IMPLEMENT THE PROFICIENCY ASSESSMENT SYSTEM AND COLLECT DATA

Steps one through seven prepare the district to be able to assess pupil proficiencies. It is now the responsibility of the school managers to implement the system as it was intended to be used. If the assessment system has weaknesses that were not apparent during development, refinements should be made as soon as possible to correct them.

Data collection should be a routine matter if the assessment system is operating as planned. It is important to the inservice staff who have data collection and reporting responsibilities that data are collected as planned and that the information is reported accurately.

Suggestion: Consider unique problems that might occur in local districts in the areas of implementing the assessment system and collecting data. Note some of these problems in the space below:

STEP 9: MAKE DECISIONS ABOUT RESPONDENDENTS BASED ON DATA

All of the assessment system planning, development and, to some extent, implementation that has occurred up to step nine could have been done in an atmosphere somewhat insulated from direct involvement with emotional reactions from people affected by the system. Once data have been collected, however, and tough decisions have to be made by pupils, parents, staff and others, strong emotional reactions can become a central concern.

The decisions that pupils must make relate basically to continuing their educational program as planned, participating in an alternative means of instruction to improve their proficiencies, and future vocational/educational choices that seem feasible for them. How these decisions are made is very important. The decisions must be arrived at in a way that guarantees retention of the pupil's integrity and feeling of self-worth.

Knowledge of assessment findings can be a powerful motivation for pupils to apply themselves to learning. Presented in a destructive manner, however, that knowledge can cause a pupil to want to avoid any future contact with school.

There are many implications for pupil personnel service staff in this area of providing feedback to pupils and helping pupils make constructive decisions based on assessment data. Psychologists and counselors, for example, may provide needed explanations to staff, parents, pupils, and others regarding the technical aspects of assessment. Such explanations should be given in language that can be understood by persons having no background in assessment. Counselors could be a resource to teaching staff regarding techniques that reduce stress and encourage positive action when teachers counsel pupils who have done poorly on the assessment and who need further instruction. Pupil personnel services staff will need a first-hand, in-depth understanding of how curriculum in their school is designed and implemented so they can better assist pupils. Perhaps they will need to receive inservice training and periodic update sessions conducted by the curriculum specialists (and/or teachers); perhaps they will need to be directly involved in curriculum design and development.

Suggestion: Consider other types of decisions that pupils must make which relate directly to assessment information, and the implications the decisions might have for staff development activities. Note some of these decisions and implications in the space below.

STEP 10 MAKE PROGRAM REFINEMENTS WHEN REQUIRED

The proficiency assessment system must help curriculum specialists refine instructional programs to make them more effective.

One fact must be remembered, however, curriculum changes. Curricula which are not responsive to changes in society will soon be of no use to pupils. New technology, better understanding of how people relate, improvements in agricultural methods, are examples of dynamics that cause curriculum change. As the curriculum changes, the assessment system must change.

Many times the first indication that the assessment procedures must change will be pupil data. These data can also point out areas where the curriculum needs refinement. For example, if the staff believed that a particular curriculum topic was being effectively implemented, but it was not, the data could show a need to improve. It must be well understood that **curriculum development and assessment are reciprocal**. When the data indicate that pupils are not learning at the desired rate, a curriculum change might also be required.

The use of assessment data to systematically make curriculum improvements is one of the most important benefits the system can provide for a district.

Suggestion: Consider possible ways assessment data can be presented to facilitate curriculum improvements. Note those procedures in the space below.

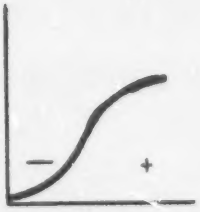
Test Validity and Reliability

In a competency-based program, achievement tests are useful only when they clearly indicate what learners have gained as a direct result of their educational experiences. This is the question of validity. **Validity** in criterion-referenced measures is established largely on the basis of a match between what is purported to be taught and what is being measured by the test items (i.e., content validity). The objectives and test specifications which guide criterion-referenced item construction provide specific delineations of acceptable student performance to demonstrate mastery. A general statement of anticipated learner outcomes often found in a "course of study" is not sufficiently specific to permit assessment.

Criterion-referenced tests must differentiate whether or not a student has mastered anticipated learnings. Scores obtained on criterion measures prior to and following instruction, therefore, should reflect change in learner competence. Since the changes being examined by criterion-referenced tests are always in relation to prespecified learning outcomes and not in relation to differences among students, correlational procedures used to establish validity for norm-referenced measures do not necessarily apply to criterion-referenced tests. Scores on norm-referenced tests are manipulated to form a normal distribution. Items which do not contribute to achieving a normal distribution are discarded or replaced. Score distributions on a criterion-referenced test, on the other hand, given prior to instruction should be negatively skewed and positively skewed when given following instruction. The differences in these distributions are shown below:

A normal distribution of scores permits comparison of students based on high or low performance.

Pre instruction distribution (positive skew)



Post instruction distribution (negative skew)

Comparisons of students to a preestablished standard can be accomplished with severely skewed distributions, but comparisons among students have little meaning.

Using a norm-referenced distribution, a correlation between relative positioning of students from pre to post-testing can give an indication of the adequacy of the test. Since the skewed distribution of a criterion-referenced test is an indication that the content being measured by the test is too difficult for the students prior to instruction, and very easy for them following instruction, the correlation between pre and post-testing will be low and may even be negative.

It is necessary for individual items in a criterion-referenced test to contribute to a separation of persons who have mastered course content from those who have not. But it is not required that items determine gradients of knowledge in a given area.

To be valid, scores on criterion-referenced measures should correlate with other measures of the objective. For example, students who are rated by teachers as "very proficient" in welding should score high on a test designed to measure welding skills.

A relatively easy way to examine the adequacy of criterion-referenced test items is to examine the percentage of students who pass items on a given test administration. The teacher may select 15 students believed to be capable of applying a math skill which has been taught. Of the 20 items used, all the students may successfully complete 13 items. Two of the items may be too difficult for eight students and five items may have been missed by three students.

The five items missed by the three students would be examined again for appropriateness of difficulty level. If the difficulty level is not too high, those three students might have been overestimated and actually were functioning at a lower level than expected.

It is also possible that some of the items which were successfully completed by all 15 students more appropriately apply to a lower level student.

Another way to test items is to administer them to a wide range of students to whom the skill has been introduced. By examining the different proportions of students who successfully respond to the item at different levels, it is possible to place the item at a level where mastery is supported by classroom performance.

Validity is the central question in criterion-referenced test construction, but unless the test is reliable it cannot be valid.

The **reliability** of a measure indicates the extent to which it is consistent in measuring whatever it is meant to measure. For example, a test that is reliable will yield the same score or information when taken more than once if the person taking the test has not changed.

Summary

Recent legal changes have placed responsibility on educators to produce students who have demonstrable proficiencies. The enormous task that this creates for educators — specifying, articulating, implementing, and assessing accountable programs — is now in sharper focus than ever. This document has provided a conceptual schema for responding to these laws, together with some exercises which illustrate possible practices educators can apply in their local settings as they implement the laws.

REFERENCES AND SUGGESTED READINGS

Adams, Georgia Sachs *Measurement and Evaluation in Education, Psychology and Guidance*. New York: Holt, Rinehart and Winston, 1974

Baker, Eva L. *Using Measurement to Improve Instruction*. September 1972. 8 pages. ED 069 762.

Before You Hand Out Those Diplomas: Should Your Students Prove They've Learned Something? American School Board Journal, 164, 3, 41-43, March 1977.

Besel, Ronald, and Okada, Masahito. *The Development of Domain-Referenced Tests for an Objectives-Based Reading Program*. April 1974. 8 pages. ED 903 918.

Biersner, Robert J. *Measurement of Reading Proficiency Among Navy Recruits*. Naval Education and Training Command, Pensacola, Fla., April 1976.

Brodinsky, Ben. *Back to the Basics: The Movement and Its Meaning*. Phi Delta Kappan, 58, 7, 522-527, March 1977.

Buffer, James John, Jr. *A Study of Certain Effects of Test Interpretation in Counseling Upon Achievement and Self-Perceptions*. Illinois School Res; 6; 2; 23-7, February 1970.

Davis, F.B. and J.J. Diamond, "The Preparation of Criterion-Referenced Tests," in Center for the Study of Evaluation Monograph Series in Evaluation No. 3. *Problems in Criterion-Referenced Measurement*, Los Angeles: University of California, 1974.

Doob, Heather S. Summary of Research on Open Education. An ERS Research Brief., Educational Research Service, Washington, D.C., 1974.

Duncan, Ann Del, Tracking Behavioral Growth: Day-to-Day Measures of Frequency Over Domains of Performance. *Educational Technology*, Vol. 14, No. 6, June 1974, pages 54-59.

Durnin, John, and Scandura, Joseph M. An Algorithmic Approach to Assessing Behavior Potential: Comparison with Item Forms and Hierarchical Technologies. *Journal of Educational Psychology*, Vol. 65, No. 2, October 1973, pages 262-272.

Educational and Occupational Choice: An Analysis of the Decision-Making Processs of Central Nebraska High School, Graduates of 1973, Final General Report, September 1975.

Ferguson, Richard L., and Tse-Chi, Hsu. *The Application of Item Generators for Individualizing Mathematics Testing and Instruction* May 1971. 21 pages. ED 053 935.

Gagne, R. M., *The Conditions of Learning*, New York: Holt, Rinehart and Winston, 1965, 1970.

Gottesman, Julia and Marian Schilling. *A Common Ground for Assessing Competencies in Written Expression* Los Angeles: Office of the Los Angeles County Superintendent of Schools, in Press.

Haladyna, Thomas. *An Analysis of Two Procedures for Decision Making When Using Domain-Referenced Tests,* April 1975, 22 pages, ED 104 957.

Haladyna, Thomas. *The Paradox of Criterion-Referenced Measurement,* April 1976, 25 pages, ED 126 155.

Haladyna, Thomas M. *The Quality of Domain-Referenced Test Items,* April 1976, 28 pages, ED 129 846.

Hall, G. R., and others. *The Evolution of Educational Performance* Contracting in Five Sch[illegible] Districts, 1971-72. A Working Note, Rand Corp., Santa Monica, CA, 1972.

Harris, Mary B., Trujillo, Amaryllis E. *Improving Study Habits of Junior High School Students Through Self-Management Versus Group Discussion*, Journal of Counseling Psychology, 22, 6, 513-511, November 1975.

Harrow, Anita J. A. *A Taxonomy of the Psychomotor Domain*. New York: David McKay Co., Inc.

Hertz, Karl V. *Learning Outside the School Building — of Course*, pager presented at National Associaton of Secondary School Principals Annual Convention (58th Atlantic City, New Jersey, March 1-6, 1974).

Hentschke, Guilbert C., and Levine, Donald M Planning for Evaluation in Performance Contracting Experiments: The Connection to Domain-Referenced Testing Theory. *Educational Technology*, Vol. 14, No. 6, pages 38-43.

Hilton, Thomas L. A Study of Intellectual Growth and Vocational Development, Final Report: Educational Testing Service, Princton, New Jersey, March 1971.

Hively, Wells. *Domain-Referenced Testing*. October 1974, 150 pages, available from Educational Technology Publications, Englewood Cliffs, New Jersey 07632.

Hively, Wells. Introduction to Domain-Referenced Testing. *Educational Technology*, Vol. 14, No. 6, June 1974, pages 5-10.

Hively, Wells, Ed., and Reynolds, Maynard C., Ed. *Domain-Referenced Testing in Special Education* 1975 146 pages. Available from Council for Exceptional Children, 1920 Associ Association Drive, Reston, Virginia 22091, Product 101.

Johnston, Thomas J. Program and Product Evaluation from a Domain-Referenced Viewpoint. *Educational Technology,* Vol. 14, No. 6, June 1974, pages 43-48.

Kosecoff, Jacquline, Fink, Arlene, and Klein, Stephen P. A System for Describing and Evaluating Criterion-Referenced Tests. Center for the Study of Evaluation, University of California, Los Angeles, CA., December 1976.

Krathwohl,David R.; Bloom, Benjamin, S.; Mesia, Bertram B. *Taxonomy of Educational Objectives - The Classification of Educational Goals, Handbook II: The Affective Domain*. New York: David McKay Co., Inc., 1964.

Kwansa, Kofi Bassa. *Investigation of the Relative Content Validity of Norm-Referenced and Domain-Referenced Arithmetic Tests.* Ph.D. Dissertation, University of Pittsburgh, 1972. (UMI Order No. 73-4153, 256 pages.)

Lembo, John M. *The Psychology of Effective Classroom Instruction*. Columbus, Ohio: Charles E. Merrill Pub. Co., 1969.

Let's ALL Work for Standards, NJEA Review, 50, 1, 20-2, September 1976.

Lewis, Paddy. Significant Others, Self-Concept of Ability and Achievement. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA., April 19-23, 1976).

Macready, George B., and Merwin, Jack C. Homogeneity Within Item Forms in Domain-Referenced Testing. *Educational and Psychological Measurement*, Vol. 33, No. 2, Summer 1973, pages 351-360.

Macready, George B. The Structure of Domain Hierarchies Found Within a Domain-Referenced Testing System. *Educational and Psychological Measurement*, Vol. 35, No. 3, Autum 1975, pages 583-597.

Macready, George Byron. An Investigation into the Nature of Interitem Relations and the Structure of Domain Hierarchies Found Within a Domain-Referenced Testing System. *Dissertation Abstracts International*, Vol. 33, No. 5-A, 1972, page 2174.

Merrill, David (ed.), *Instructional Design: Readings*, Englewood Cliffs: Prentice-Hall, 1971.

Millman, Jason, Criterion-Referenced Measurement. In Popham, W.J. *Evaluation in Education*. Berkeley: McCutchan Pub. Corp., 1974.

Millman, Jason. Sampling Plans for Domain-Referenced Tests. *Educational Technology*, Vol. 14, No. 6, June 1974, pages 17-21.

Nitko, A.J., "Problems in the Development of Criterion-Referenced Test: The IPI Pittsburg Experience," in Center for the Study of Evaluation Monograph Series in Evaluation No. 3. *Problems in Criterion-Referenced Measurement*, Los Angeles: University of California, 1974.

Nitko, Anthony J. *Some Considerations When Using a Domain-Referenced System of Achievement Tests in Instructional Situations.* March 2, 1970. 24 pages, ED 037 793.

Nygaard, Debra D. *Evaluating of Year-Round School Programs*. An ERS Research Brief. Educational Research Service, Washington, D.C., 1974.

Olympia, P.L., Jr. *Repetitive Domain-Referenced Testing Using Computers: The TITA System*. June 1975, 9 pages, ED 111 358.

O'Reilly, Robert P., and others. *The Validation and Refinement of Measures of Literal Comprehension in Reading for Use in Policy Research and Classroom Management.* February 1976, 424 pages, ED 133 363.

Peter, L.J., *Competencies for Teaching Individual Instruction,* Belmont, California: Wadsworth, 1972, 1975.

Petersen, Sonya, Magaro, Peter A. *Reading and Field Dependence:* A Pilot Study. J Reading: 12, 4, 287-94, January 1969.

Plakos, John et al, *Developing Useful Objectives,* Sacramento: California Evaluation Improvement Project, 1975.

Popham, James W. Teacher Evaluation and Domain-Referenced Measurement. *Educational Technology*, Vol. 14, No. 6, June 1974, pages 35-37.

Popham, W. J., *Educational Evaluation*, Englewood Cliffs: Prentice-Hall, 1974, Chapter 7.

Popham, W. James. *An Evaluation Guidebook: A Set of Practical Guidelines for the Educational Evaluator*. 89 pages. The Instructional Objectives Exchange, 1972.

Popham, W. James. *"Observations on the Hart Law (AB 3408)"* Los Angeles: Instructional Objectives Exchange, November 8, 1976.

Pupil Progression Plan for the Dade County Public Schools, Miami, Fla., 1977.

Rasumssen, L. V. *Individualizing Science Education*. January 1970.

Sanders, James R., and Murray, Stephen L. Alternatives for Achievement Testing. *Educational Technology*, Vol. 16, No. 3, 7 pages, March 1976.

Saylor, J. Galen. *The School of the Future — Now*. Association for Supervision and Curriculum Development, Washington, D.C., 1972.

Sension, Donald B., and Rabehl, George. Test-Item Domains and Instructional Accountability. *Educational Technology,* Vol. 14, No. 6, pages 22-28, June 1974.

Simpson, Elizabeth, "The Classification of Educational Objectives, Psychomotor Domain." *Illinois Teacher of Home Economics*. 1966-67.

Skager, R.W., "Generating Criterion-Referenced Tests from Objectives-Based Assessment Systems: Unsolved Problems in the Development, Assembly, and Interpretation," in Center for the Study of Evaluation Monograph Series in Evaluation No. 3. *Problems in Criterion-Referenced Measurement,* Los Angeles: University of California, 1974.

Smith, Robert L., and Troth, William A. *Journal of Counseling Psychology,* November 1975.

Thayer, J.A. *Johnny Could Read — What Happened?* J Reading; 13; 7; 501-06, 561, April 1970.

Vitola, Bart M., "Item Writers Handbook: Techniques for Test Construction," Monterey, California: CTB/McGraw-Hill, n/d, mimeograph.

Wiersma, William and Jars, S. G., *Evaluation of Instruction in Individually Guided Education Reading*, Massachusetts: Addison-Wesley, 1976, p.224.

Willoughby, Lee, and others. *A Comparison of Domain-Referenced and Classic Psychometric Test Construction Methods*. 13 pages. ED 131 128, 1976.

Whitely, Susan E. Domain-Referenced Testing: An Alternative Model for Test Construction. *Proceedings of the Annual Convention of the American Psychological Association*, Vol. 6, Pt. 2, pages 515-516, 1971.

An Official Publication
Office of the Los Angeles County Superintendent of Schools,
Division of Program Evaluation, Research and Pupil Services

Richard M. Clowes, *Superintendent*

Los Angeles County Board of Education

Robert M. Bock, *President*
Dr. Daniel L. Towler, *Vice-President*
Lawrence J. Kaplan, D.D.S.
Mrs. Louise H. Marsh
Dr. Earl V. Pullias
Vincent H. Simpson
Mrs. Kathryn Vanderhook

September 1977